CHARACTERIZATION OF SOYBEAN SEED YIELD USING OPTIMIZED PHENOTYPING

by

BRENT SCOTT CHRISTENSON

B.A., Saint John's University, 2011

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
William T. Schapaugh Jr.

Copyright

Abstract

Crops research moving forward faces many challenges to improve crop performance. In breeding programs, phenotyping has time and economic constraints requiring new phenotyping techniques to be developed to improve selection efficiency and increase germplasm entering the pipeline. The objectives of these studies were to examine the changes in spectral reflectance with soybean breeding from 1923 to 2010, evaluate band regions most significantly contributing to yield estimation, evaluate spectral reflectance data for yield estimation modeling across environments and growth stages and to evaluate the usefulness of spectral data as an optimized phenotyping technique in breeding programs. Twenty maturity group III (MGIII) and twenty maturity group IV (MGIV) soybeans, arranged in a randomized complete block design, were grown in Manhattan, KS in 2011 and 2012. Spectral reflectance data were collected over the growing season in a total of six irrigated and water- stressed environments. Partial least squares and multiple linear regression were used for spectral variable selection and yield estimation model building. Significant differences were found between genotypes for yield and spectral reflectance data, with the visible (VI) having greater differences between genotypes than the near-infrared (NIR). This study found significant correlations with year of release (YOR) in the VI and NIR portions of the spectra, with newer released cultivars tending to have lower reflectance in the VI and high reflectance in the NIR. Spectral reflectance data accounted for a large portion of variability for seed yield between genotypes using the red edge and NIR portions of the spectra. Irrigated environments tended to explain a larger portion of seed yield variability than water-stressed environments. Growth stages most useful for yield estimation was highly dependent upon the environment as well as maturity group. This study found that spectral reflectance data is a good candidate for exploration into optimized phenotyping techniques and with further research and validation datasets, may be a suitable indirect selection technique for breeding programs.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Precision Phenotyping Using Canopy Reflectance Measurements in Field Crops Research: A Review

Introduction

One of the challenges of plant breeders moving forward is to create genotypes that produce more with fewer inputs and on fewer acres. Over the last decade, advancements in genotyping techniques have reached new heights, whereas phenotyping techniques have stayed stagnant (Furbank and Tester, 2011). As crops research moves forward, characterization and screening of large populations as well as more efficiently selecting superior genotypes for quantitative traits such as yield is necessary. The phenotyping constraints make this exploration and selection process difficult due to the time and economic investment as well as the destructive nature and single time point of such techniques (Reynolds et al., 1999). To increase efficiency in crop research and development and utilize the genotypic resources available, precise and effective phenotyping techniques need to be developed (Montes et al., 2007; Furbank, 2009).

These challenges and constraints have led to studies focused on precision phenotyping techniques using canopy reflectance and canopy temperature measurements that are both precise and high-throughput allowing for more genetic material to be screened in less amount of time (Reynolds et al., 2007). These techniques look to accurately characterize morphological and physiological traits that can be used as an indirect selection method and improve yield (Richards et al., 2001; 2002). Precision phenotyping techniques can also give researchers a repeatable, accurate, nondestructive method of characterizing plant functions throughout the growing season (Hatfield et al., 2008).

Precision phenotyping is based on the reflectance and absorption characteristics of vegetation and the spectral response recorded by remote sensors (Holland et al., 2006). Remote

sensors have been utilized in close proximity with the plant, mounted on tractors for high-throughput measurements, or mounted on satellites, such as LandSat. These spectral responses can then be used to calculate vegetation indices or in the case of full spectra techniques, used independently in model building. Most of the indices, such as the widely used NDVI, are functions of ratios or proportions between wavelengths that correlate to specific plant biophysical or biochemical properties such as leaf composition, biomass production, or pigmentation (Osborne et al., 2002).

In a research setting, the accurate prediction of quantitative traits such as yield can be used as an important tool for identifying superior lines or used as a screening tool in many trait mining studies, without the need for high input costs such as harvesting (Ma et al., 2001; Montes et al., 2007). This can lead to larger breeding programs that utilize more plots and material in the pipeline. To accurately estimate yield, an understanding of the biophysical/biochemical components influencing the spectra to understand what the sensor is sensing as well as conclusions made based upon the spectral response curves. These biophysical/biochemical components, such as canopy and leaf structure and pigment status may not allow the plant to fully reach its genetic yield potential (Clevers, 1997), and many researchers have suggested that yield gains seen in field crops can be attributed to more efficient photosynthetic parameters (Waddington et al., 1986; Caldirini et al., 1995; Sayre et al., 1997; Reynolds et al., 1999).

Plant Biophysical and Biochemical Properties

Leaf reflectance and emittance of solar radiation is the basis for canopy reflectance and canopy temperature research (Inman et al., 2007). The amount of reflectance observed is a product of the leaf tissues, cellular structure, and the air-cell wall-protoplast-chloroplast interaction (Kumar and Silva, 1973). Chlorophyll a and b as well as anthocyanins and

carotenoids preferentially reflect and absorb wavelengths, which can then be used to estimate the concentration and status of these pigments (Hatfield et al., 2008). The wavelengths that correlate well with certain plant functions or parameters can then be used to model plant phenotypes (Thomas and Oerther, 1972; Filella et al., 1995). Emitted and reflected wavelengths can be used in the same fashion to correspond to canopy temperature using infrared thermometers (Gausman et al., 1969; Blackmer et al., 1994).

Canopy reflectance Properties

Canopy structure has a tremendous influence on the electromagnetic spectral response obtained by remote sensors. Because researchers are collecting solar radiation reflecting from the leaf surface, canopy biochemical and biophysical properties play a large role in what light and in what capacity the light is in, that the sensors are detecting (Asner et al., 1998). As the canopy structure changes, light may be reflected or diffused from different portions of the leaf and the canopy, causing a magnitude of variation from sample-to-sample (Asner et al., 1998). Also, soil background may confound these scans, creating unwanted variability in the sample. However, researchers can also take advantage of these variations and model these differences (Richie et al., 2010). By modeling the canopy structure, researchers can make inferences about the health of the plant and desired attributes such as photosynthetic capabilities (Thomas and Gausman 1977; Wessman, 1990).

Early research was focused on finding new wavelengths and spectral regions that correlated to plant function. Tucker (1978) proposed 5 primary and 2 transition regions of the visible and near infrared spectrum to characterize plant functions. 400-500nm (blue region) correlates to chlorophyll and carotenoid concentrations, 500-620 (green region) correlates to a reduction in chlorophyll absorption, which is why plants appear green, 620-700nm (red) which

correlates to an upswing in chlorophyll absorption (creation of the red edge), 700-740, which is a transition stage from red absorption to near infrared radiation (NIR) reflection, resulting in a sharp spike in reflection percentages in the spectral response, and from 750-1100 nm. The last region, NIR, is a high reflection region and also contains possible water absorption regions. Signal to noise ratio in remote sensing research is always a concern and sensing reflectance of plant canopies increases this ratio (Weber et al., 2012). Equations that have been proposed try to take into account background aspect such as soil and dead biomass with some success, but researchers have yet to fully account for all in-field variables when capturing reflectance data from full plant canopies (Daughtry et al., 2000; Hatfield et al., 2008).

Water Status

Evaluating water status using any technique has its basis in the fundamentals of water status and movement in the plant, such as total water content, water potential of the roots and cells, and transpiration and photosynthetic rates (Pearcy et al., 1989). Water stress research using remote sensing techniques has been used for many years, starting with Tanner (1963) studying plant temperature variations from air temperature (Hatfield et al., 2008). This early research led to characterizing water stress through leaf and canopy temperature and water absorption bands within the EMS. This research has even been used to create irrigation schedules (Bausch and Duke, 1996) to reduce excess water use and characterize superior water use efficient plants in breeding programs.

Researchers have been able to characterize plant water status and stress through the preferential absorption of water and more specifically the hydroxyl ions within the electromagnetic spectrum (Peñuelas et al., 1993; 1997; Gao, 1996; Serrano et al., 2000). The NIR (730-1300) and middle infrared (1300-2500 nm) has been shown to correlate well with

water status/content of plants and the wavelengths 970, 1240, 1400, and 2700 nm have been proposed as water absorption bands indicative of water status (Tucker, 1980; Peñuelas et al., 1993; Gao, 1996; Zarco-Tejada et al., 2001; Gutierrez et al., 2010).

Peñuelas et al. 1993 developed the water index (WI) defined as R_{970}/R_{900}, to predict water stress and was found to strongly correlate with the relative water content of the plants and. The normalized difference water index, proposed by Gao (1996) defined as $(R_{860}-R_{1240}/R_{860}+R_{1240})$ has been used to evaluate water stress in corn and soybeans (Anderson et al., 2004) with favorable results. Babar et al., 2006a proposed normalizing the WI, using the wavelengths 970, 900, and 850nm, developed by Peñuelas et al., 1993 found it to be useful for screening wheat genotypes for water stress and water use efficiency (Gutierrez et al., 2010). Prasad et al., 2007a, b also normalized the WI, using the wavelengths 970, 920, and 880nm to screen winter wheat lines under dry-land conditions (Gutierrez et al., 2010). These authors propose that the normalization of the WI provides added genotypic variation explanation. However, Gutierrez et., al 2010 found that only the normalized water index using wavelengths 970 and 880 nm proposed by Prasad et al., 2007 a, b was the only index sufficient for predicting water stress within field experiments.

Leaf Biochemical Properties

Leaf biochemistry is a major portion of the physiological function of plant leaves and serves as the building block for energy production. Many of the physiological parameters that show genotypic variation pertain to biochemical parameters, with most of the components being pigment concentration and status (Reynolds et al., 2009). The photosynthetic capacity and efficiency of a plant is a function of the chlorophyll and other pigment content of the plant.

Chlorophyll status can then estimate biomass production and photosynthetic capacity (Curran et al., 1990; Filella et al., 1995). Also, due to chlorophyll content being directly influenced by nitrogen status, chlorophyll status estimation can directly relate to nitrogen status (Filella et al., 1995; Moran et al., 2000). Chlorophyll content is also related to plant stress and senescence, suggesting chlorophyll can be used to characterize genotypes for stress (Hendry et al., 1987; Merzlyak and Gitelson, 1995; Peñuelas and Filella, 1998; Merzlyak et al., 1999; Carter and Knapp, 2001).Remote sensing leaf pigments can serve as a non-destructive, repeatable way to characterize plant functions and relate this to overall plant health (Hatfield and Prueger, 2010).

With new advancements in the understanding of leaf structure and remote sensing interaction, many modern studies have developed indices expressing the relationship between leaf biochemical properties and spectral responses obtained. Peñuelas et al., 1995 and Gitelson and Merzlyak, 1994 have expresses various chlorophyll and chlorophyll:carotenoid ratios using specially derived indices, Gamon et al., 1992, 1997 have characterized photosynthetic performance based on spectral responses. The ratio $(R_{531}-R_{570})/(R_{531}+R_{570})$, where R_{570} is the reflectance at 570nm and R_{531} is the reflectance at 531 nm, has been widely used to characterize leaf pigment status (Gamon et al., 1992; Peñuelas et al., 1995). Chappelle et al., 1992 found that wavebands 650, 675, and 700 nm could be used to predict chlorophyll and beta carotene in soybean leaves. Also, alternative wavelengths have been used based on correlation to specific physiological functions (Gamon et al., 1992; Inoue et al., 2006; Garbulsky et al., 2008). Being able to model genotypic responses that incorporate these functions have potential to exploit important traits in breeding programs.

Chlorophyll A and B are essential pigments within a plant that convert light energy into chemical energy through the process of photosynthesis. From chlorophyll, researchers can

directly determine photosynthetic potential and primary productivity of the plant (Curran et al., 1990; Filella, 1995). Chlorophyll A and B absorb solar radiation at specific wavelengths to covert this energy into chemical energy. Based on the wavelengths that are either absorbed or reflected by these pigments, scientists are able to assess overall health and efficiency of the plant. Chlorophylls have a high absorbance in the red (600-730 nm) and blue (400-500nm) regions of the spectrum. Chlorophyll b preferentially absorbs light in the 460 and 650 nm regions, chlorophyll a absorbs in the 580, 630, and 670 nm regions of the spectrum, and 415 nm is a preferential absorption of both (Chappelle et al., 1992). Blue regions of the spectrum also correlate to carotenoid preferential absorption, and therefore are not commonly used in chlorophyll concentration and status estimates (Sims and Gamon, 2002). The green and red region of the spectrum around the 550nm and 700nm regions are primarily used due to high chlorophyll concentrations needed to saturate these wavelengths (Sims and Gamon, 2002). Many researchers such as Thomas and Gausman, 1977; Buschman and Nagel, 1993; Datt, 1998, 1999; Gitelson and Merzylak 1996; 1997; Gitelson et al., 1996b; and Schepers et al., 1996, have characterized the chlorophyll wavelength absorption into chlorophyll concentration models utilizing mainly the red. Sims & Gamon, (2002), correlated the 700nm region of the spectrum to chlorophyll concentration, which is used by many researchers as a base line for new chlorophyll indices. Gitelson and Merzlyak 1994 used (R_{780}/R_{700}) and (R_{750}-R_{705}/R_{750}+R_{705}), where R_{780} is the reflectance at 780nm, R_{700} is the reflectance at 700nm, R_{750} is the reflectance at 750nm and R_{705} is the reflectance at 705nm, to estimate chlorophyll content with great success. Other researchers have also used the 550nm and 700 nm regions of the spectrum (Aoki et al., 1986; Gitelson and Merzlyak, 1997; Datt, 1998; Gamon and Surfus, 1999; Carter and Knapp, 2001; Richardson et al., 2002; Sims and Gamon, 2002). Chlorophyll content has many variables, and

Gitelson et al. 2002 suggest that broadband reflectance is necessary for broad scale adaptation of chlorophyll estimation across species and ecosystems.

Anthocyanins and carotenoids are important pigments of the plant biochemical process as well. Few researchers have delved into anthocyanin and carotenoid research due to numerous variables such as other pigment absorption and leaf cell structure problems. Gamon and Surfus (1999), developed a model and indices for anthocyanin concentration estimation based on a red to green ratio, R_{600}-R_{700}/R_{500}-R_{600}, and the green channel on LandSat has been explored by Vina and Gitelson et al., 2011, to estimate anthocyanin status with success, but has yet to be replicated. Hatfield et al., 2008 hypothesized that 2 band indices are confounded by variables such as other pigment absorption and leaf scattering of light. Gitelson et al., 2001, Gitelson et al., 2003a, and Gitelson et al., 2006, found 3-band indices in the (waveband regions) ineffective in developing reliable models to estimate carotenoid and anthocyanin concentrations, suggesting full spectrum hyperspectral data collection may be necessary for reliable model estimations.

Many authors have studied chlorophyll and other pigment models with varying success, but the consensus seems to be, full spectra indices or statistical models are necessary to fully capture the relationship between leaf pigments and reflectance spectra. However, models utilizing regions know to correlate well with plant function can be used to characterize many traits that are of importance to crop researchers. These models may also be able to be used for screening of specific traits that may not be necessarily captured by visual phenotyping techniques.

Yield Estimation

With the new challenges of feeding an ever-growing population with fewer resources, it is necessary for crops researchers to streamline the breeding approach and develop techniques

that allow for more genetic material to be screened within a program in a shorter amount of time. Due to high genetic variability and extensive genotype x environment interactions associated with yield, precise phenotyping is essential for future development. Most of the field based research that is conducted on yield estimation models using canopy reflectance and canopy temperature measurements are focused on 2 or 3 band indices, which can be highly variable and inconsistent (Babar et al., 2006 a,b). Some of these indices have been useful in estimating yield through other plant characteristics such as chlorosis (Adams et al., 1999), green cover (Dusek et al., 1985; Daughtry et al., 2000), chlorophyll (Datt, 1999; Daughtry et al., 2000), and photosynthetically active tissue (Wiegand et al., 1991). These indices, like pigment status have had varying degrees of success, but none have fully captured the underlying physiological and environmental factors leading to phenotypic yield. New studies that focus on utilizing full spectrum instruments and models have been published in recent years in wheat (Hansen et al., 2002; Hansen and Schjoering, 2003; Pimstein et al., 2007; 2011), corn (Hong et al., 2001; Weber et al., 2012), rice (Lin et al., 2012), cotton (Zhao et al., 2006) and soybean (Kaul et al., 2005) in optimal and drought environments. Incorporating these new models into high efficiency platforms has also been characterized and explored (Montes et al., 2007; Walter et al., 2012; White et al., 2012). Other researchers have also related canopy temperature, which is emitted thermal radiation to yield (Reynolds et al., 1994; Fischer et al., 1998). Combining reflectance models with canopy temperature can integrate many physiological parameters that lead to robust yield prediction models that account for a large portion of the variability among genotypes (White et al., 2012). Leaf and canopy reflectance data can give insight into many physiological parameters such as photosynthetic capacity, aboveground biomass, and water status that have highly correlated association with yield (Royo et al., 2003; Weber et al., 2012). Many of these

studies have focused on wheat and are seen to correlate more effectively with optimal environments due to yield potential being fully met (Aparicio et al., 2000; Royo et al., 2003; Gutieirrez et al., 2010).

Canopy Temperature for Yield

Canopy temperature (CT) measurements are a fast and accurate way to study stomatal conductance and leaf transpiration (Jones et al., 2009). Canopy temperature measurements are based on using thermal infrared (IR) radiation which relates to evaporation or transpiration from a plant leaf. When a plant stops transpiration and closes the stomata, the temperature of the leaf and ultimately canopy increases. This is due to the assumption that transpired water evaporated from the leaf surface and cools the leaf below ambient air temperatures (Jackson, 1981). CT measurements can then be used to distinguish genotypes and make selection for high yielding varieties (Reynolds et al., 2009).

Canopy temperature and canopy temperature depression (CT - ambient air temperature) can be used for many essential plant functions. Early research focused on crop stress indicators using canopy temperature with great success (Jackson et al., 1981). Blum et al., 1989 related yield stability to canopy temperatures in wheat under drought conditions with correlations of 0.64 and 0.72 between the drought susceptible index and CT. In soybean Fletcher et al., 2007 and Ries et al., 2012 found that canopy temperature can be utilized to characterize slow wilting and radiation use efficiency. Research has suggested that lower CT correlate with high yielding wheat varieties in well watered and stressed environments (Reynolds et al., 1999; Babar et al., 2006b; Gutierrez et al., 2010). Research also suggests that selecting for lower CT varieties can increase yields in a breeding program in well watered and water stressed environments (Amani et al., 1996; Reynolds et al., 2009). Research conducted suggests a correlation with CT and grain

yield in soybean (McKinney et al., 1989), cotton (Hatfield et al., 1987), millet (Singh and Kanemasu, 1983), as well as wheat. Incorporating canopy temperature into breeding programs for high yielding varieties may be a useful indirect selection tool for plant breeders. Therefore, CT is well suited to be utilized in yield estimation models.

Reflectance indices for Yield

Recent research conducted suggests that reflectance wavelength equations called indices correlating to yield can be used as an indirect selection method (Araus et al., 2001). Most of the research conducted on yield prediction using canopy reflectance techniques focuses on the reflected solar radiation in the visible red (600nm-730nm) and near infrared (730nm-1100nm) regions of the electromagnetic spectrum, which was first shown to correlate to crop conditions by Bauer 1981 and Walburg et al., 1982. Researchers can then develop rapid, nondestructive, and repeatable measurements (Field et al., 1995).

Vegetation indices are used to maximize the relationship between certain solar radiation wavelengths and plant function, while minimizing the effect of background noise (Huete et al. 2002; Hatfield and Prueger 2010). Hatfield and Prueger 2010 studied various vegetation indices in corn and soybean and concluded that multiple indices must be used to account for the variations seen between different crops and different growing conditions. Babar et al., 2006a concluded that vegetation indices using NIR reflectance were the most correlated with yield in wheat. Wiegand et al., 1991 also illustrates that yield prediction by spectral indices must be site-independent in order to create a prediction model that can be used and fully accepted in crop research.

Most of the indices correlate plant parameters such as pigment status to grain yield. Indices such as the simple ratio (SR), first used by Jordan 1969 and Rouse et al., 1973 described

253 by the equation (R_{NIR}/R_{RED}), captures the ratio of NIR reflectance to reflectance in the red. SR has
254 been shown to correlate well with biomass, LAI, fractional photosynthetically active radiation
255 (FPAR), and ground cover, and yield in wheat and soybeans (Hatfield, 1983; Wiegand et al.,
256 1991; Ball and Konzak, 1993; Price and Bausch 1995; Aparicio et al., 2000; Serrano et al., 2000;
257 Ma et al., 2001; Royo et al., 2003; Hatfield and Prueger, 2010). The basis for the SR is the strong
258 absorption of red light and the strong reflection of NIR by healthy plant vegetation, and small
259 changes in these reflectance or absorption patterns can give researchers a lot of information
260 about overall plant health, especially chlorophyll a and b. The normalized difference vegetation
261 index (NDVI), derived by Deering 1978 and Tucker 1979 to estimate green biomass and
262 intercepted PAR, defined as (R_{NIR}-R_{RED}/R_{NIR}+R_{RED}), has been used extensively to predict yield and
263 other plant functions with many crops using hyper-spectral and satellite imagery (Wiegand et al.,
264 1991; Peñuelas et al., 1997; Lewis et al., 1998; Aparicio et al., 2000; Ma et al., 2001; Shanahan
265 et al., 2001; Royo et al., 2002; Royo et al., 2003; Prasad et al., 2007a;b; Marti et al., 2007). The
266 NDVI captures plant health, with healthy plants having high absorption in the red and high
267 reflection in the NIR. If the plant is under stress or is starting to senesce, the absorption of red
268 light by the chlorophylls and other pigments and cellular components will decrease. Gitelson et
269 al., 1996a proposed the GNDVI as a substitute to the high saturation point of the red region in
270 the NDVI. Shanahan et al., 2001 have successfully predicted yields in corn under normal
271 growing conditions in Nebraska using the GNDVI. Researchers found that normalizing the green
272 and NIR relationship was highly correlated with grain yield, explaining 70 to 92% of yield
273 variability at mid grain fill in corn (Shanahan et al., 2001). Ma et al., 2001 used the NDVI (R_{613}-
274 R_{559}/R_{613}+R_{559}) to predict soybean yield, and concluded a high correlation with soybean yields
275 under irrigated conditions and explained up to 80% of the variation within yield. The

276 photochemical reflectance index (PRI), defined by the equation $(R_{550} - R_{531})/(R_{550} + R_{531})$
277 captures the normalized difference between the major green wavelength reflectance of the plant
278 canopy and leaves, which can be used to quantify radiation use efficiency (Gamon et al., 1992).
279 Trotter et al., 2002 and Garbulsky 2011 used the PRI to assess nitrogen use efficiency and
280 radiation use efficiency to distinguish genotypes that were superior compared to checks. These
281 indices have been used estimate yield varying environments on many platforms.
282 Other indices focus on the biophysical parameters, such as water content and leaf area.
283 The leaf area index (LAI), first illustrated by Tucker and Sellers, 1986 was developed to predict
284 vegetation parameters such as green biomass and green leaf area (Babar et al. 2006b). The leaf
285 area index captures the genotypic differences between photosynthetically active radiation (PAR)
286 being absorbed and used by leaves with more area and being able to monitor important crop
287 variables throughout the growing season (Clevers, 1997). Elliot and Regan 1993 and Aparicio et
288 al., 2000 and 2002 have determined that LAI plays a large part in plant function and correlates
289 well with yield prediction. Many water indices have been developed to relate reflectance
290 measurements with success in the 850 and 970 nm regions with bread and durum wheat yield
291 (Royo et al., 2003; Prasad et al., 2007b; Gutierrez et al., 2010). The water indices focus on the
292 minor water absorption bands within the spectrum and capture relative water content based on
293 absorption strength by water within the plant leaves. Capturing biophysical properties can
294 capture properties that influence yield production of field crops.
295 Most of the research conducted using reflectance measurements for yield estimation
296 focus on calculated indices. These indices are easy to calculate and give researchers a way to
297 easily handle the large amount of data associated with such research, however, indices tend to be
298 environment specific and outside of wheat, have a low correlation with yield (Aparicio et al.,

2000; Ma et al., 2001; Guitierrez-Rodrigues et al., 2004; Prasad et al., 2007a and b; Gutierrez et al., 2010). Therefore, more complex models need to be developed using more of the spectrum to capture the necessary variation within yield across different environments and crops (Osborne et al., 2002; Pimstein et al., 2007; 2011; Weber et al., 2012).

Yield estimation using hyper spectral data

Recently, studies utilizing full spectrum models focusing on several hundred bands contributing to yield estimation have been conducted. Full spectrum models are considered for yield estimation due to confounds such as biomass saturation, dense cover, high LAI, and high chlorophyll levels (Pimstein et al., 2007). High yielding varieties tend to have high biophysical/biochemical properties, making it unreliable to distinguish these genotypes using spectral indices (Baret and Guyot, 1991; Buschmann and Nagel, 1993; Aparicio et al., 2002; Pimstein et al., 2007). None of the developed indices can meet these needs, therefore, more complex characterization models based on more of the spectra must be developed (Hansen and Schjoering, 2003). Due to yield being a function of many physiological and environmental parameters that change throughout the growing season, combining many different wavebands over an entire growing season should give more precise yield estimation models (Araus et al., 2008).

Full spectrum multivariate data analysis with spectral reflectance data to create yield estimation models in field crops has been utilized in recent years (Hong et al., 2001; Hansen et al., 2002; Chang et al., 2003; Ferri et al., 2004; Kaul et al., 2005; Poss et al., 2006; Pimstein et al., 2007; 2011; Gutierrez et al., 2011; Weber et al., 2012; Lin et al., 2012). This is mainly due to the large datasets associated with spectral research, high multicollinearity that is associated with wavebands close to each other, and fears of over fitting models with more bands (Thenkebail et

al., 2002; Lin et al., 2012). To deal with these dilemmas, researchers have turned to approaches such as partial least squares (PLS), principal component analysis (PCA), and artificial neural networks (ANN) to develop hyper spectral yield estimation models. Incorporating more bands into a model may allow for more variability within certain predictors without reducing the effectiveness of the model.

The main approach used to deal with the challenges of hyper spectral data analysis focus on reducing the correlation between predictor variables, causing multicollinearity and normalizing the variability within predictor variables from sample to sample. Also, researchers use these new techniques to tease apart the spectral response curves into meaningful spectral waveband regions. PCA has been used by researchers to build yield prediction models that explain 70 to 90% of the yield variability in maize (Hong et al., 2001; Chang et al., 2003) and 39% of the variability in yield among soybean genotypes (Hong et al., 2001). Stepwise regression explained almost 95% of the variability within maize yield using 6 bands (Osborne et al., 2002). ANN has also been used in soybean to build prediction models that explain between 46 and 81% of the variability in yield and 42 to 77% in corn (Kaul et al., 2005). Hansen et al., 2002 compared three different PLS methods and concluded the N-PLS was the most consistent for yield and protein content estimation in wheat and barley, explaining up to 75% of the variation in protein content and up to 97% of the variation in yield. Similarly, under different water regimes, Weber et al., 2012 found that PLS explained a maximum of 40% of the variability within corn yield, and found that prediction models created explained more variability in water limited environments than non-water limited environments. Lin et al., 2012 used orthogonal projections to latent structure PLS in rice and could distinguish 3 cultivars with 90% accuracy but dropped to 80% when distinguishing 1 cultivar. Researchers also determined that

reducing the dataset to 10 nm intervals reduced the noise within the dataset. Utilizing multivariate analysis is new to precision phenotyping research but appears to have potential to create yield prediction models as well as explain other important phenotypic parameters.

Conclusion

With the challenges needing to be met by plant breeders moving forward, new and innovative phenotyping technologies need to be developed that allows for more genetic material and advanced genotyping technologies to be utilized. Precision phenotyping using canopy reflectance data can be utilized by plant breeders to characterize a large number of genotypes in a precise and cost effective manner. Reflectance indices can be used to classify many different plant functions that are vital to characterizing field crops and screen genotypes for traits vital for plant breeders to meet the challenges. With the advancements in statistical analysis, hyper spectral data can be used to create models that can quantify quantitative traits such as yield and allow plant breeders to screen more genotypes in more environments. Moving forward, more research needs to be conducted in order for yield prediction models to be fully implemented into breeding programs, but reflectance data can be utilized to characterize large genotype experiments.

References

Adams, M.L., Philpot, W.D., and Norvell, W.A. 1999. Yellowness index: an application of spectral 2nd derivatives to estimate chlorosis of leaves in stressed vegetation. International Journal of Remote Sensing. 20:3663–3675.

Amani, I., Fischer, R.A., and Reynolds, M.P. 1996. Canopy temperature depression association with yield of irrigated wheat cultivars in a hot climate. J. Agron. Crop Sci. 176:119-129.

Anderson, M.C., C.M.U. Neale, F. Li, J.M. Norman, W.P. Kustas, H. Jayanthi, and J. Chavez. 2004. Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. Remote Sens of Env. 92: 447-464.

Aoki, M., Yabuki, K., Totsuka, T., and Nishida, M. 1986. Remote sensing of chlorophyll content of leaf: I. Effective spectral reflection characteristics of leaf for evaluation of chlorophyll content in leaves of dicotyledons. Environmental Control in Biology. 24: 21– 26.

Aparicio, N., Villegas, D., Araus, J., Casadesus, J., and Royo, C. 2002. Relationship between growth traits and spectral vegetation indices in durum wheat. Crop Science. 42(5):1547-1555.

Aparicio, N., Villegas, D., Casadesus, J., Araus, J., and Royo, C. 2000. Spectral vegetation indices as nondestructive tools for determining durum wheat yield. Agronomy Journal. 92(1):83-91.

Araus, J.L., J. Casadesus, and J. Bort. 2001. Recent tools for the screening of physiological traits determining yield. p. 59–77. In M.P. Reynolds, J.I. Ortiz-Monasterio, and A. Mcnab (ed.) Application of physiology in wheat breeding. CIMMYT, Mexico, D.F.

Asner, G.P., Wessman, C.A., and Archer, S., 1998. Scale dependence of absorption of photosynthetically active radiation in terrestrial ecosystems. Ecol. Appl. 8 (4):1003–1021.

Babar, M., Reynolds, M., Van Ginkel, M., Klatt, A., Raun, W., and Stone, M. 2006a. Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. Crop Science. 46(2):578-588.

Babar, M., Reynolds, M., Van Ginkel, M., Klatt, A., Raun, W., and Stone, M. 2006b. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. Crop Science. 46(3):1046-1057.

Ball, S., and Konzak, C. 1993. Relationship between grain yield and remotely sensed data in wheat breeding experiments-. Plant Breeding. 110(4):277-282.

Baret, F., and Guyot, G. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment, Remote Sens. Environ. 35:161-174.

Bauer, M. E., C. S. T. Daughtry and V. C. Vanderbilt. 1981. Spectral agronimic relationships of corn, soybean, and wheat canopies. Report SR-P1-04187. West Lafayette, IN: Laboratory for Applications of Remote Sensing, Purdue University. pp: 17.

Bausch, W.C., and H.R. Duke. 1996. Remote sensing of plant nitrogen status in corn. Trans. ASAE. 32:1869–1875.

Blackmer, T.M., J.S. Schepers, and G.E. Varvel. 1994. Light reflectance compared with other nitrogen stress measurements in corn leaves. Agron. J. 86:934–938.

Blum, A. 1988. Breeding for stress environments. CRC Press, Boca Raton, FL.

Buschman, C., and Nagel, E. 1993. In vivo spectroscopy and internal optics of leaves as a basis for remote sensing of vegetation. International Journal of Remote Sensing. 14:711 – 722.

Calderini, D.F., M.F. Dreccer, and G.A. Slafer. 1995. Genetic improvement in wheat yield and associated traits. A re-examination of previous results and the latest trends. Plant Breeding. 114:108–112.

Carter, G.A., and A.K. Knapp. 2001. Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. Am. J. Bot. 84:677–684.

Chang, J., Clay, D.E., Dalsted, K., Clay, S., O'Neill, M. 2003. Corn (*Zea Mays* L.) yield prediction using multispectral and multidate reflectance. Agron. J. 95:1447-1453.

Chappelle, E.W., M.S. Kim, and J.E. McMurtrey, III. 1992. Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. Remote Sens. Environ. 39:239–247.

Clevers, J.G.P.W. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. Remote Sens. Environ. 61:221–228.

Curran, P.J., J.L. Dungan, and H.L. Gholz. 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. Tree Physiol. 7:33–48.

Datt, B. 1999. Remote sensing of water content in Eucalyptus leaves. Australian Journal of Botany. 47: 909 - 923.

Datt, B. 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. Remote Sens. Environ. 66:111–121.

Daughtry, C.S.T., C.L. Walthall, M.S. Kim, E. Brown de Colstoun, and J.E. McMurtrey, III. 2000. Estimating corn leaf chlorophyll content from leaf and canopy reflectance. Remote Sens. Environ. 74:229–239.

Deering, D.W. 1978. Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. Ph.D. diss. Texas A&M Univ., College Station.

Dusek, D.A., Jackson, R.D. and Musick, J.T. 1985. Winter wheat vegetation indices calculated from combinations of seven spectral bands. Remote Sensing of Environment. 18:255–267.

Elliot, G., and Regan, K. 1993. Use of reflectance measurements to estimate early cereal biomass production on sandplain soils. Australian Journal of Experimental Agriculture. 33:179-183.

Ferri, C.P., Formaggio, A.R., Schiavinato, M.A. 2004. Narrow band spectral indexes for chlorophyll determination in soybean canopies [Glycine max (L.) Merril]. Brazilian Journal of Plant Physiology. 16: 131–136.

Field, C.B., Randerson, J.T., Malmström, C.M. 1995. Global net primary production: combining ecology and remote sensing. Remote Sens. Environ. 51(1):74–88.

Filella, I., I. Serrano, J. Serra, and J. Penuelas. 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminate analysis. Crop Sci. 35:1400–1405.

Fischer, R.A., D. Rees, K.D. Sayre, Z.M. Lu, A.G. Condon and A. Larque Saavedra. 1998. Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. Crop Sci. 38:1467–1475.

Fletcher, A.L., T.R. Sinclair, and L.H. Allen, Jr. 2007. Transpiration responses to vapor pressure deficit in well watered "slow-wilting" and commercial soybean. Environ. Exp. Bot. 61:145–151.

Furbank R.T., Tester, M. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. Trends in Plant Science. 16 (12):635-644.

Furbank, R.T. 2009. Plant phenomics: from gene to form and function. Funct. Plant Biol. 36:10–11.

Gamon, J. A., Serrano, L., and Surfus, J. S. 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types and nutrient levels. Oecologia. 112:492– 501.

Gamon, J.A., and J.S. Surfus. 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143:105–117.

Gamon, J.A., J. Penuelas, and C.B. Field. 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41:35–44.

Gao, B. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment. 58:257– 266.

Garbulsky, M. 2011. The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies. Remote Sensing of Environment. 115:281-297.

Garbulsky, M.F., Penuelas, J., Papale, D., and Filella, I. 2008. Remote estimation of carbon dioxide uptake by a Mediterranean forest. Global Change Bio. 14:2860-2867.

Gausman, H.W., W.A. Allen, and R. Cardenas. 1969. Reflectance of cotton leaves and their structure. Remote Sens. Environ. 1:19–22.

Gitelson, A.A., and Merzlyak, M.N. 1997. Remote estimation of chlorophyll content in higher plant leaves. International Journal of Remote Sensing. 18: 2691–2697.

Gitelson, A.A., and M.N. Merzlyak. 1996. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. J. Plant Physiol. 148:494–500.

Gitelson, A., Gritz, Y., and Merzlyak, M. 2003a. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology. 160:271-282.

Gitelson, A., Kaufman, Y., and Merzlyak, M. 1996a. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment. 58:289-298.

Gitelson, A., M. Merzlyak, and H. Lichtenthaler. 1996b. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. J. Plant Physiol. 148:501–508.

Gitelson, A., Zur, Y., Chivkunova, O., & Merzlyak, M. 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochemistry and Photobiology. 75:272-281.

Gitelson, A.A., A. Via, T.J. Arkebauer, D.C. Rundquist, G. Keydan, and B. Leavitt. 2003b. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30:1248.

Gitelson, A.A., and M.N. Merzlyak. 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J. Photochem. Photobiol. 22:247–252.

Gitelson, A.A., G.P. Keydan, and M.N. Merzlyak. 2006. Tree-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophys. Res. Lett. 33:111402 .

Gitelson, A.A., M.N. Merzlyak, and O.B. Chivkunova. 2001. Optical properties and non-destructive estimation of anthocyanin content in plant leaves. Photochem.Photobiol. 74:38–45.

Gitelson, A.A., Thenkabail, P.S., Lyon, J.G., Huete, A. 2011. Remote Sensing estimation of crop biophysical characteristics at various scales. Chapter 15 in Hyperspectral Remote Sensing of Vegetation. 329-358, Taylor and Francis.

Gutierrez, M., Reynolds, M., Raun, W., Stone, M., and Klatt, A. 2010. Spectral water indices for assessing yield in elite bread wheat genotypes under well-irrigated, water-stressed, and high-temperature conditions. Crop Science. 50:197-214.

Gutierrez-Rodriguez, M., Reynolds, M.P., Escalante-Estrada, J.A., Rodriguez-Gonzalez, M.T. 2004. Association between canopy reflectance indices and yield and physiological traits in bread wheat under drought and well-irrigated conditions. Australian Journal of Agricultural Research 55:1139–1147.

Hansen, P.M., and Schjoering, J.K. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens. Environ. 86(4):542-553.

Hansen, P.M., Jorgenson, J.R., Thomsen, A. 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squared regression. J. of Ag Sci. 139:307-318.

Hatfield, J. L., Gitelson, A.A., Schepers, J.S., and Walthall, C. L. 2008. Application of Spectral Remote Sensing for Agronomic Decisions. Papers in Natural Resources. Paper 257.

Hatfield, J., & Prueger, J. 2010. Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. Remote Sensing, 2:562-578.

Hatfield, J.L. 1983. Remote sensing estimators of potential and actual crop yield. Remote Sens. Environ. 13:301–311.

Hatfield, J.L., J.E. Quisenberry, and R.E. Dilbeck. 1987. Use of canopy temperatures to identify water conservation in cotton germplasm. Crop Sci. 27:269–273.

Hendry, G.A.F., J.D. Houghton, and S.B. Brown. 1987. The degradation of chlorophyll-a biological enigma. New Phytol. 107:255–302.

Holland, K.H., J.S. Schepers, and J.F. Shanahan. 2006. Configurable multispectral active sensor for high-speed plant canopy assessment. In D.J.

Hong, S.Y., K.A. Sudduth, N.R. Kitchen, H.L. Palm, and W.J. Weibold. 2001. Using hyperspectral remote sensing data to quantify within-field spatial variability [CD ROM]. Proc. Int. Conf. on Geospatial Inf. in Agric. and Forestry, 3rd, Denver, CO. 5–7 Nov. 2001. Altarum, Ann Arbor, MI.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Of Environment, 83:195-213.

Inman, D., Khosla, R., Reich, R., and Westfall, D. 2007. Active remote sensing and grain yield in irrigated maize. Precision Agriculture. 8:241-252.

Inoue, Y., and Peñuelas, J. 2006. Relationship between light use efficiency and photochemical reflectance index in soybean leaves as affected by soil water content. International J. of Remote Sens, 27:5109−5114.

Jackson, R.D., S.B. Idso, R.J. Reginato, and P.J. Pinter, Jr. 1981. Canopy temperature as a crop water stress indicator. Water Resour. Res. 17:1133–1138.

Jones, H.G., R. Serraj, B.R. Loveys, L. Xiong, A. Wheaton, A.H. Price. 2009. Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Functional Plant Biol. 36:978-989.

Jordan, C.F. 1969. Derivation of leaf area index from quality of light on the forest floor. Ecology 50:663–666.

Kaul, M., R.L. Hill, C. Walthall. 2005. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems 85:1-18.

Kumar, R., and L. Silva. 1973. Light ray tracing through a leaf cross-section. Appl. Optics 12:2950–2954.

Lewis, J.E., Rowland, J., Nadeau, A., 1998. Estimating maize production in Kenya using NDVI: some statistical considerations. Int. J. Remote Sens. 13, 2609–2617.

Lin, W.S., C.M. Yang, B.J. Kuo. 2012. Classifying cultivars of rice (*Oryza sativa* L.) based on corrected canopy reflectance spectra data using the orthogonal projections of latent structures (O-PLS) method. Chemometrics and Intelligent Laboratory Systems, 115:25-36.

Ma, B.L., Dwyer, L.M., Costa, C., Cober, E.R., Morrison, M.J. 2001. Early Prediction of Soybean yield from canopy reflectance measurements. Agron. J. 93:1227-1234.

Marti, J., Bort, J., Slafer, G., & Araus, J. 2007. Can wheat yield be assessed by early measurements of normalized difference vegetation index? Annals of Applied Biology. 150:253-257.

McKinney, N.V., Schapaugh, W.T., Kanemasu, E.T. 1989. Canopy temperature, seed yield, and vapor pressure deficit relationships in soybean. Crop Sci. 29:1038-1041.

Merzlyak M.N., Gitelson A.A. 1995. Why and what for the leaves are yellow in autumn? On the interpretation of optical spectra of senescing leaves (*Acer platanoides* L.). J Plant Physiol 145: 315–320.

Merzlyak MN, Gitelson AA, Chivkunova OB, Rakitin VY. 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum 106:135–141.

Montes, J.M., Melchinger, A.E., Reif, J.C. 2007. Novel throughput phenotyping platforms in plant genetic studies. Trends in Plant Sci. 12(10):433-436.

Moran, J.A., A.K. Mitchell, G. Goodmanson, and K.A. Stockburger. 2000. Differentiation among effects of nitrogen fertilization treatments on conifer seedlings by foliar reflectance: A comparison of methods. Tree Physiol. 20:1113–1120.

Osborne, S., Schepers, J., Francis, D., and Schlemmer, M. 2002. Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and water-stressed corn. Crop Science, 42:165-171.

641 Pearcy, R.W., Ehleringer, J.R., Mooney, H.A., Rundel P.W. 1989. Measurement of transpiration
642 and leaf conductance. In Pearcy et al. Plant Physiological Ecology: Field methods and
643 instrumentation. 137-160.
644
645 Peñuelas, J., & Filella, I. 1998. Visible and near-infrared reflectance techniques for diagnosing
646 plant physiological status. Trends in Plant Science, 3:151-156.

647 Peñuelas, J., F. Baret, and I. Filella. 1995. Semi-empirical indices to assess
648 carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31:221–230.
649
650 Peñuelas, J., Filella, I., Biel, C., Serrano, L., & Save, R. 1993. The reflectance at the 950– 970
651 mm region as an indicator of plant water status. International Journal of Remote Sensing, 14,
652 1887–1905.
653
654 Peñuelas, J., R. Isla, I. Filella, and J.L. Araus. 1997. Visible and near infrared reflectance
655 assessment of salinity effects on barley. Crop Sci. 35:1400–1405. Sci. 37:198–202.
656
657 Pimstein, A., Karnieli, A., & Bonfil, D. 2007. Wheat and maize monitoring based on ground
658 spectral measurements and multivariate data analysis. Journal of Applied Remote Sensing.
659 1:013530.

660 Pimstein, A., Karnieli, A., Bansal, S., & Bonfil, D. 2011. Exploring remotely sensed
661 technologies for monitoring wheat potassium and phosphorus using field spectroscopy. Field
662 Crops Research. 121:125-135.

663 Poss, J., Russell, W., & Grieve, C. 2006. Estimating yields of salt- and water-stressed forages
664 with remote sensing in the visible and near infrared. Journal of Environmental Quality. 35:1060-
665 1071.

666 Prasad, B., Carver, B., Stone, M., Babar, M., Raun, W., & Klatt, A. 2007a. Genetic analysis of
667 indirect selection for winter wheat grain yield using spectral reflectance indices. Crop Science.
668 47:1416-1425.

669 Prasad, B., Carver, B., Stone, M., Babar, M., Raun, W., & Klatt, A. 2007b. Potential use of
670 spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains
671 conditions. Crop Science. 47: 1426-1440.

672 Price, J.C., and W.C. Bausch., 1995. Leaf area index estimation from visible and near-infrared
673 reflectance data. Remote Sens of Environ. 52:55–65.
674
675 Reynolds, M., Dreccer, F., and Trethowan, R. 2007. Drought-adaptive traits derived from wheat
676 wild relatives and landraces. J. Exp. Bot. 58: 177–186.
677
678 Reynolds, M., Manes, Y., Izanloo, A., Langridge, P. 2009. Phenotyping approaches for
679 physiological breeding and gene discovery in wheat. Annals of Appl. Bio. 155(3):309-320.
680

Reynolds, M.P., M. Balota, M.I.B. Delgado, I. Amani and R.A. Fischer. 1994. Physiological and morphological traits associated with spring wheat yield under hot, irrigated conditions. Aust. J. Plant Physiol. 21:717–730.

Reynolds, M.P., S. Rajaram and K.D. Sayre. 1999. Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci. 39:1611–1621.

Richards, R. A., G. J. Rebetzke, A. G. Condon, and G. D. Farquhar. 2002. Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. Crop Sci. 42: 111-121.

Richards, R.A., A. G. Condon, and G.J. Rebetzke. 2001. Traits to improve yield in dry environments. P. 88-100. In Application of physiology in wheat breeding. CIMMYT, Mexico.

Richardson, A.D., S.P. Duigan, and G.P. Berlyn. 2002. An evaluation of noninvasive methods to estimate foliar chlorophyll content. New Phytol. 153:185–194.

Ries, L.L., L.C. Purcell, T.E. Carter Jr., J.T. Edwards, C.A. King. 2012. Physiologial traits contributing to differential canopy wilting in soybean under drought. Crop Sci. 52: 272-281.

Ritchie, G., Sullivan, D., Vencill, W., Bednarz, C., & Hook, J. 2010. Sensitivities of normalized difference vegetation index and a Green/Red ratio index to cotton ground cover fraction. Crop Science, 50:1000-1010.

Rouse, J.W., Jr., R.H. Haas, J.A. Schell, and D.W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. p. 309–317. In Proc. Earth Res. Tech. Satellite-1 Symp., Goddard Space Flight Cent., Washington, DC. 10–14 Dec. 1973.

Royo, C., Aparicio, N., Villegas, D., Casadesus, J., Monneveux, P., & Araus, J. 2003. Usefulness of spectral reflectance indices as durum wheat yield predictors under contrasting Mediterranean conditions. International Journal of Remote Sensing. 24:4403-4419.

Sayre, K.D., S. Rajaram, and R.A. Fischer. 1997. Yield potential progress in short bread wheat in Northern Mexico. Crop Sci. 37: 36–42.

Schepers, J.S., T.M. Blackmer, W.W. Wilhelm, and M. Resende. 1996. Transmittance and reflectance measurements of corn leaves from plants with diff erent nitrogen and water supply. J. Plant Physiol. 148:523–529.

Serrano, L., I. Filella, and J. Peñuelas. 2000. Remote sensing of biomass and yield of winter wheat under diff erent nitrogen supplies. Crop Sci. 40:723–731.

Shanahan, J.F., J.S. Schepers, D.D. Francis, G.E. Varvel, W.W. Wilhelm, J.S. Tringe, M.R. Schlemmer, and D.J. Major. 2001. Use of remote sensing imagery to estimate corn grain yield. Agron. J. 93:583–589.

Sims, D.A., and J.A. Gamon. 2002. Relationship between leaf pigment content and spectral refl ectance across a wide range species, leaf structures and development stages. Remote Sens. Environ. 81:337–354.

Tanner, C.B. 1963. Plant temperature. Agron. J. 55:210–211.

Thenkabail PS, Smith RB, De Pauw E. 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. Photogrammetric Engineering and Remote Sensing 68: 607–621.

Thomas, J.R., and G.F. Oerther. 1972. Estimating nitrogen content of sweet pepper leaves by reflectance measurements. Agron. J. 64:11–13.

Thomas, J.R., and H.W. Gausman. 1977. Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops. Agron. J. 69:799–802.

Trotter, G. M., Whitehead, D., & Pinkney, E. J. 2002. The photochemical reflectance index as a measure of photosynthetic light use efficiency for plants with varying foliar nitrogen contents. International J. of Remote Sens. 23:1207−1212.

Tucker, C. J. 1980. Remote sensing of leaf water content in the near infrared. Remote Sensing of Environment, 10:23– 32.

Tucker, C. J. and Sellers, P. J. 1986. Satellite remote sensing of primary production. International Journal of Remote Sensing. 7:395-1416.

Tucker, C. J., 1978. A comparison of satellite sensors for monitoring vegetation. Photogr. E. R. 44: 1369-1380.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8:127–150.

Vina, A., & Gitelson, A. 2011. Sensitivity to foliar anthocyanin content of vegetation indices using green reflectance. IEEE Geoscience and Remote Sensing Letters. 8:464-468.

Waddington, S.R., J.K. Ransom, M. Osmanzai, and D.A. Saunders. 1986. Improvement in the yield potential of bread wheat adapted to Northwest Mexico. Crop Sci. 26:698–703.

Walburg, G., M.E. Bauer, C.S.T. Daughtry, and T.L. Housley. 1982. Light-use efficiency of a winter wheat crop subjected to N and water deficiencies. Agron. J. 74:677–683.

Walter, A., B. Studer, R. Ko¨lliker. 2012. Advanced phenotyping offers opportunities for imporved breeding of forage and turf species. Annals of Botany. 1-9.

Weber, V.S., J.L. Araus, J.E. Cairns. C. Sanchez, A.E. Melchinger, E. Orsini. 2012. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. Field Crops Researcch. 128: 82-90.

Wessman, C. A. 1990. Evaluation of canopy biochemistry. In Remote Sensing of Biosphere Functioning. Eds R. J. Hobbs & H. A. Mooney. 135–156.

White, J.W., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, M.M. Conley, K.A. Feldmann, A.N. French, J.T. Heun, D.J. Hunsake, M.A. Jenks, B.A. Kimball, R.L. Roth, R.J. Strand, K.R. Thorp, G.W. Wall, G. Wang. 2012. Field-Based phenomics for plant genetics research. Field Crops Research. 133:101-112.

Wiegand, C., Richardson, A.,Escobar, D., and Gerbermann, A. 1991. Vegetation indexes in crop assessment. Remote Sensing of Environment. 35:105-119.

Zarco-Tejada, P. J., Miller, J. R., Noland,T. L., Mohammed, G. H., and Sampson, P. H. 2001. Scaling-up and model inversion methods with narrow-band optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. IEEE Transactions on Geosciences and Remote Sensing. 39:1491–1507.

Zhao, D., K.R. Reddy, V.G. Kakani, J.J. Read, S. Koti. 2007. Canopy reflectance in cotton for growth assessment and lint yield prediction. Europ. J. Agronomy, 26:335-344.

Chapter 2 - Characterizing Changes in Soybean Spectral Response Curves with Breeding Advancements

Abstract

Soybean (*Glycine max* (L.) Merr.) crop yield has steadily increased in the past 60 years due in part to breeding advances. Spectral reflectance correlated to specific plant functions may help characterize the impact of breeding on soybean cultivar development. The objectives of this study were:1) to find specific regions of the soybean spectra response curves that show genotypic differences; and 2) to determine the effect of the breeding process on spectral response curves of soybean cultivars. Spectral reflectance measurements were taken on 20 maturity group III (MGIII) and 20 maturity group IV (MGIV) soybean genotypes ranging in release year from 1923 to 2010 (arranged in a randomized complete block design) in 2011 and 2012 in Manhattan, KS. Significant genotypic differences were found between entries, especially in the green (500 nm– 600 nm), red (600 nm–700 nm), and red-edge (700nm–730 nm) portions of the spectra. This study also found significant correlations with year of release (YOR) in the visible (VIS) and near-infrared (NIR) spectra. The newer released cultivars tended to have lower reflectance values in the VIS and red-edge spectra portions and higher values in the NIR portion of the spectra than older cultivars. This study concluded that breeding has most affected spectral reflectance curves by reducing the VIS portion of the spectra and extending the red-edge, resulting in a shift to lower reflectance values further into the NIR and then a sharp inflection to the NIR, with higher values in the NIR. In addition, a crossover occurs in new genotypes around 1150 nm, resulting in lower reflectance values in the transition from the NIR to the middle infrared. These results suggest that breeding

advancement has had an impact on spectral reflectance curves and the areas that have been changed with breeding advancement may be exploited for further advancement.

Introduction

One way to characterize plant phenotypes and function is by measuring spectral reflectance. The amount of reflectance observed is a product of the leaf tissues, cellular structure, and air-cell wall–protoplast-chloroplast interaction (Kumar and Silva, 1973). The visible (VIS) portion of the spectra (400–730 nm) have low reflectance due to the absorptive attributes of chlorophyll and other accessory pigments such as carotenoids, carotenes, and anthocyanins; however, the near-infrared portion (NIR) has a high reflectance caused by the scattering of light by cellular components and water. Many of the physiological parameters that show genotypic variation — most of which are pigment concentration and status — pertain to biochemical parameters (Reynolds et al., 2009). The photosynthetic capacity and efficiency of a plant is a function of the chlorophyll and other pigment content of the plant. Chlorophyll status can also be used to estimate biomass production and photosynthetic capacity (Curran et al., 1990; Filella et al., 1995), and, because chlorophyll content is directly influenced by nitrogen, chlorophyll status estimation can relate directly to nitrogen status (Filella et al., 1995; Moran et al., 2000). By modeling the leaf structure and content, researchers can make inferences about the plant's health and desired attributes, such as photosynthetic capabilities and water status (Thomas and Gausman, 1977; Wessman, 1990). Researchers also have estimated yield through other plant characteristics such as chlorosis (Adams et al., 1999), green cover (Dusek et al., 1985; Daughtry et al., 2000), chlorophyll (Datt, 1999; Daughtry et al., 2000), photosynthetically active tissue (Wiegand et al., 1991), and water status (Prasad et al., 2007a, 2007b).

Chlorophylls have a high absorbance in the red (600–700 nm) and blue (400–500 nm) regions of the spectrum. Chlorophyll b preferentially absorbs light in the 460 and 650 nm regions, chlorophyll a absorbs in the 580, 630, and 670 nm regions of the spectrum, and 415 nm is a preferential absorption of both in soybean (Chappelle et al., 1992). Blue spectrum regions also correlate to carotenoid preferential absorption and therefore are not commonly used in chlorophyll concentration and status estimates (Sims and Gamon, 2002). The green and red regions of the spectrum—around 550 nm and 700 nm, respectively—are primarily used, due to high chlorophyll concentrations needed to saturate these wavelengths (Sims and Gamon, 2002). Shanahan et al. (2001) explained from 70 to 92% of yield variability at mid-grain fill in corn using the green region and NIR. Sims and Gamon (2002) correlated the 700 nm region to chlorophyll concentration, which is used by many researchers as a baseline for new chlorophyll indices. The ratio $(R_{531}-R_{570})/(R_{531}+R_{570})$ has been widely used to characterize leaf pigment status in many species (Gamon et al., 1992; Peñuelas et al., 1995). Chappelle et al. (1992) found that wavebands 650, 675, and 700 nm could be used to predict chlorophyll ($R^2 = 0.934$) and beta carotene ($R^2 = 0.935$) in soybean leaves. Gitelson and Merzlyak (1994) used (R_{780}/R_{700}) and $(R_{750}-R_{705}/R_{750}+R_{705})$ to explain as much as 98% of the variability within chlorophyll content in chestnut and maple trees. The normalized difference vegetation index (NDVI) derived by Deering (1978) and Tucker (1979) to estimate green biomass and intercepted PAR, defined as $(R_{NIR}-R_{RED}/R_{NIR}+R_{RED})$, has been used to predict yield and other plant functions with many crops using hyperspectral and satellite imagery (Wiegand et al., 1991; Peñuelas et al. 1997; Ma et al. 2001; Shanahan et al. 2001; Royo et al. 2002; Royo et al. 2003; Prasad et al. 2007a, 2007b; Marti et al. 2007). Ma et al. (2001) explained up to 80% of the yield variability in soybean using 613 and 813 nm. Others have also used the 550 and 700 nm regions of the spectrum when leaves are

yellow-green (Gitelson and Merzlyak, 1996; Datt 1998; Gamon and Surfus, 1999; Sims and Gamon, 2002) to estimate chlorophyll status and yield, but high absorption in these regions may lead to saturation at relatively low chlorophyll concentrations. Gamon and Surfus (1999) developed a model and indices for anthocyanin concentration estimation based on a red-to-green ratio, R_{600}-R_{700}/R_{500}-R_{600}, and Viña and Gitelson (2011) used broadband wavelengths to estimate anthocyanin status (R^2 = 0.25 – 0.89) with peak absorption around 540 to 560 nm.

Plant water status and stress through the preferential absorption of water and, more specifically, the hydroxyl ions within the electromagnetic spectrum also can be used to characterize plants (Peñuelas et al., 1993; Peñuelas et al., 1997; Gao, 1996; Serrano et al., 2000). The NIR (700–1300) and middle infrared (1300–2500 nm) ranges have been shown to correlate well with water status or content of plants and the wavelengths 970, 1240, 1400, and 2700 nm have been proposed as water absorption bands (Peñuelas et al., 1993; Gao, 1996; Gutierrez et al., 2010). Babar et al. (2006a, 2006b) used the wavelengths developed by Peñuelas et al. (1993) and found them useful for screening wheat genotypes for water stress and water-use efficiency. Prasad et al. (2007a, 2007b) used the wavelengths 970, 920, 900, 880, and 850 nm to screen winter wheat lines under dryland conditions and found significant correlations with yield. Gao (1996) normalized wavebands (R_{860}-R_{1240}/R_{860}+R_{1240}) and evaluated water stress in corn and soybean, with favorable results.

Spectral reflectance correlated to specific plant functions may help characterize the impact of breeding on soybean cultivar development. The objectives of this study were:1) to find specific regions of the soybean spectral response curves that show genotypic differences and could be exploited for genotype distinction; and 2) to determine the effect of the breeding process on spectral response curves of soybean cultivars released from 1923 to 2010.

Materials and Methods

Experimental and Field Design

A study was conducted on soybean [*Glycine max*. (L) Merr.] at the Kansas State University Research farm south of Manhattan, KS, in 2011 and 2012. In 2011, experiments (location A) were conducted on a well-drained silt loam soil of the Eudora and Belvue types. In 2012, location B was a well-drained silt loam soil of the Bismarckgrove-Kimo complex, and location C was a well-drained silt-sandy loam soil of the Belvue and Eudora types.

Twenty maturity group III (MGIII) and 20 maturity group IV (MGIV) soybean cultivars, ranging in release year from 1923 to 2010 (Table 1), were selected for release-year diversity out of the Soybean Genetic Gain Study coordinated by Dr. Brian Diers, University of Illinois. Selected cultivars were a mixture of private and public genotypes.

Replicated plots were planted on 23 May 2011 on field A, and 16 May 2012 and 4 June 2012 on fields B and C, respectively. Each experimental unit consisted of 4 rows (3.4 meter long; spaced 76 cm apart). Genotypes were planted in separate well-watered and water-stressed environments arranged in a randomized complete block design with four replications. In both years, weed pressure and fertility were not limiting. Flood irrigation was applied to the well-watered environments starting at reproductive stage 1 (R1) and continued weekly until R6. In 2011, no supplemental irrigation was applied to the water-stressed environments. Due to extremely dry conditions, irrigation was applied once shortly after R1 in 2012 to ensure crop development in the water-stressed environments.

Phenotypic Traits

Maturity, height, and lodging were taken on all plots during both seasons. Maturity was calculated as the number of days past 31 August when 95% of the pods had reached mature plant

color. Height (cm) was measured as the distance from the base of the plant to the top of the main stem. Lodging was scored on a scale of 1 to 5, based on the amount of leaning or broken plants. Upright plants with no lean were scored as 1; other scores were 2 = 20° lean, 3 = 45° lean, 4 = 60° lean, and 5 = flat on the ground. The center two rows of each plot were mechanically harvested using a 2-row plot combine. Seed yield was recorded as kg ha^{-1}, adjusted to uniform moisture.

Canopy Reflectance Measurements

Canopy reflectance measurements were conducted using an ASD FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, CO). Solar radiation reflecting back from the plant canopy was captured from 350 to 2500 nm in the electromagnetic spectrum using fiber optics with a 25° field of view and sampling intervals of 1.4 nm between 350 and 1050 nm and 2 nm between 1050 and 2500 nm. Moving averages were calculated automatically to achieve 1-nm-width continuous bands. A white calibration disk ($BaSO_4$) was used to achieve reflectance percentages and to calibrate the spectroradiometer with a dark current and white reference every 40 plots or when needed (ranging from every plot to 40 plot intervals, depending on field conditions). The sensor was mounted on an adjustable monopod pole and held vertically 1 m above the canopy to achieve a 50-cm circumference collection area. Two measurements were taken per plot on rows 2 and 3, excluding the first meter of each plot to eliminate border effect. Each measurement was the average of 10 scans, which was calculated automatically. Spectral data were collected on nearly cloud-free days within ±2 hours of solar noon. Measurements were collected weekly from R3–R6 in 2011 and R2–R6 in 2012 (Fehr and Caviness, 1977), totaling three and four collection dates for the water-stressed and well-watered environments in 2011, respectively. In 2012, there were five collection dates for the well-watered MGIV environment

on field B and four collection dates for the well-watered MGIII experiment on field B and well-watered MGIII and MGIV experiments on field C. There were three collection dates for the water-stressed experiment on field B, but only two collection dates for the water-stressed environment on field C for each maturity group.

Data Pretreatment

Hyperspectral data were initially trimmed from 350–2500 nm to 400–1310 nm, which was necessary to eliminate noise from atmospheric absorption regions focused around the water absorption bands in the infrared and atmospheric scatter of blue color in the ultraviolet portion of the spectrum. Before averages were calculated for observation day and season totals, initial outlier control was implemented on each observation day's raw data. After outliers were identified and excluded from analysis, data were combined to form 10-nm-wide band regions to reduce the dataset size and eliminate some of the collinearity associated with bands in close proximity (Naes et al., 2004). Combined bands have been determined to contain less variation from sample to sample than single-band measurements (Lin et al., 2012).

Statistical Analyses

All data were analyzed using SAS 9.2 (SAS Institute, 2008). The GLIMMIX procedure of SAS (SAS Institute, 2008) was used for analysis of variance (ANOVA) of yield. Cultivar and experiment were treated as fixed effects and replication nested within the experiment was treated as a random effect. PROC GLM was used for ANOVA of the wavebands. Spectral data were analyzed for significant difference between genotypes using the genotype × environment interaction as the error term. PROC CORR was used to characterize the relationship between YOR and each waveband.

Results and Discussion

Genotypic Performance for Yield and Reflectance Data

Genotypes differed significantly in average seed yield across the six environments in both maturity groups (Table 2.1). Seed yield ranged from 2.32 to 4.63 t ha^{-1} and 1.97 to 4.26 t ha^{-1} for the MGIII and MGIV experiments, respectively. Seed yield increased with year of release by 0.0238 and 0.0255 t ha^{-1} $year^{-1}$ and had coefficient of determination values of 82 and 86% between yield and YOR in MGIII and MGIV, respectively (Figure 2.1). Yield increases are consistent with observations of short-season soybean covering 58 years of cultivar development (Morrison et al., 1999). Significant differences between genotypes, environments, and genotype × environment interaction were observed for seed yield ($P < 0.0001$) in both maturity groups (Table 2.2). The genotype × environment interactions were quite low compared with the total phenotypic variation of each maturity group experiment because the genotypes tested were genetically diverse with a large gradient between high-yielding and low-yielding genotypes.

Significant differences ($P < 0.05$) were detected between genotypes for individual band regions in both maturity groups, with the exception of 905 nm in MGIII (Table 2.3). In the MGIII experiment, Illini was eliminated from spectral analysis due to spectral inconsistencies from lodging and soil confounds. The VIS portion of the spectrum (405–695 nm) had greater genotypic differences than the NIR (705–1305 nm) in both maturity groups. The highest genotypic differences (based on F values) were observed in the green and red portions of the spectrum in both maturity groups. In the MGIII and MGIV experiments, the 735 to 1135 nm region was the least significant portion of the spectrum. Large G × E interactions and high variability in the NIR from cellular scatter can account for the decreased genotypic differences observed for this region. The MGIV experiment had higher genotypic differences observed than

995 the MGIII experiment because G × E interactions constituted less of the total phenotypic
996 variation.

997 Spectral responses varied based on YOR in both maturity groups. The earliest released
998 cultivars tended to have higher values in the VIS portion of the spectrum and lower values in the
999 NIR than the latest released cultivars (Figures 2.2 and 2.3). Maturity group spectral curves were
1000 similar in magnitude. High values for group VIS suggests that chlorophyll and other associated
1001 pigments are not as plentiful or efficient in the earlier cultivars, and higher values in the green
1002 correlates with more yellow pigment and therefore more foliar diseases or issues contributing to
1003 decreased plant function. These results are consistent with trends observed in soybean (Ma et al.,
1004 2001), corn (Chang et al., 2003; Weber et al., 2012), wheat (Hansen et al., 2002), and rice (Lin et
1005 al., 2012). Higher reflectance in the NIR suggests more biomass production or denser canopy
1006 and cellular structure in the later cultivars (Sims and Gamon, 2003; Royo et al., 2003). In the
1007 MGIV experiment, greater separation was observed between cultivars in the 735– 1095 nm—an
1008 area correlated with brown pigmentation, biomass production, and water status—than in the
1009 MGIII experiment (Figure 2.3). Observed results suggest that foliar diseases and maturity play a
1010 larger role in MGIV than in MGIII. The MGIII experiment genotypes separate much more in the
1011 1155–1305 nm regions (Figure 2.2), which have been associated with biomass production, water
1012 content, and water status. In both experiments, genotypes tended to be stable in the VIS and NIR:
1013 entries exhibited high reflection in the VIS but lower NIR reflection values. Weber et al. (2012)
1014 found a crossover in the water content region and concluded that lower values in this region
1015 correlated with high-yielding corn cultivars. A water-content region crossover was also observed
1016 in this experiment, which validates Weber et al.'s assertion for soybean and suggests that the
1017 crossover region may be a candidate for genotypic differentiation.

Correlation of Spectral Reflectance Data to Year of Release

Many of the genotypic differences in spectral reflectance were significantly correlated (r) to YOR (Figure 2.4). Compared with the NIR portion of the spectrum, highest r values were observed in the VIS portion in both experiments. The green (505-595 nm), red (605–695 nm), and red-edge (705–735 nm) regions of spectrum (r = -0.79 to -0.83, r = -0.76 to -0.85, and r = -0.76 to -0.88 respectively) exhibited the highest correlations with YOR in both MG genotypes (Figure 4). The NIR was not as consistent between maturity groups: the 1155–1305 nm had significant association with YOR in the MGIII experiment and the 765–1125 nm regions in the MGIV experiment.

For the MGIII experiment, r values ranging from -0.88 to 0.6 were observed for waveband to YOR correlation. For the MGIV experiment, r values ranged from -0.93 to 0.60. The green and red regions correlate with the green reflection and red absorption by chlorophyll and other photosynthetic pigments. The blue portion of the spectrum (405–495) also had significant correlation with YOR (r = -0.61 to -0.77) in the MGIII and MGIV (r = -0.53 to -0.93) because of absorption peaks in chlorophyll a and b and beta carotene within soybean leaves in this region (Chappelle et al., 1992). Wavebands in the blue also have been associated with canopy temperature measurements. Canopy temperature has been shown to have a significantly negative correlation with soybean YOR and yield. The blue region in field-based research tends to be unreliable, however, due to atmospheric scatter of blue light; hence, r = -0.93 in the MGIV experiment seems like an unlikely result. Abnormally high reflectance values were obtained in the 405 waveband region from some of the earlier-released cultivars. This result may be due to atmospheric scatter or the tendency for older cultivars to lodge and allow soil confounds to influence the spectra more.

The red inflection region (705–735 nm) was highly significantly correlated with YOR in the MGIII (r = 0.65–0.84) and MGIV (r = 0.79–0.88) experiments (Figure 2.4). In the MGIII experiment, the 705-nm region had the highest correlation with year, but in the MGIV experiment, the 715-nm region had the most significant correlation. Filella and Peñuelas (1994) characterized the red-edge inflection point as the transition region between red and NIR, where a sharp increase in reflection values are observed. They found that the region correlated well with overall leaf health, in which the magnitude of the difference between the reflection values between the red and NIR is a function of chlorophyll content and health, plant nitrogen status, and plant water status. Plants with a steeper increase in reflection values between the two regions have high chlorophyll content, high leaf area, and overall healthier vegetation. Weber et al. (2012) found that the red-edge region contributed significantly to yield estimation, which explains a significant portion of yield variation in corn cultivars. This study found negative correlations between YOR and reflection values in this region. Compared with other cultivars, a decrease in reflection is associated with a shift of the red edge to longer wavelengths, which occurs because of increased chlorophyll content and increased biomass production (Filella and Peñuelas, 1994).

Near-infrared measurements were not as well correlated with YOR as the VIS spectrum, but significant correlations were found between YOR and the NIR plateau (745–895 nm), as well as the small water absorption region (915–1065 nm) in the MGIV entries. Decreased spectral reflectance in the plateau region correlates with brown pigment (senescence), and significant correlations (r = 0.32 to 0.60) suggest that maturity and a higher incident of foliar diseases occurred in the later (rather than the earlier) MGIV cultivars. The maturity factor may be due to the extremely hot and dry conditions in 2011 and 2012 and an early frost in 2012, which

prevented some MGIV cultivars from reaching full maturity. Visual observations indicated that earlier-released MGIV cultivars tended to have more foliar issues leading to early senescence and less green tissue due to insects and disease than earlier-released MGIII genotypes. The plateau region has been shown to be associated with total biomass production, which suggests that total biomass may have increased with YOR. As in the plateau region, the small water absorption region had a positive relationship between YOR, and these band regions may suggest that total water content has increased as breeders have made selections based on other characteristics.

This experiment found significant correlations with YOR in the 1100–1305-nm regions in the MGIII entries but not in the MGIV genotypes. The 1100–1305-nm regions have been associated with plant water status with high absorption and lower reflection of light by water leading to greater water content within the plant leaf (Sims and Gamon, 2003). In this experiment, later- released genotypes had a higher reflection in these regions than earlier-released genotypes. Sims and Gamon (2003) concluded that thick canopies (such as those seen in the newer-released cultivars) tend to confound water absorption regions, leading to increased reflection values. The greater reflectance values would be expected because less light penetrates thicker vegetation than thinner vegetation (Bull, 1991). Asrar et al. (1983) found that NIR and red reflectance could be used for leaf area index (LAI) calculation (which measures photosynthetically active tissue) and concluded that high reflectance in the NIR was associated with higher LAI values in wheat. However, LAI values are highly influenced by soil backgrounds, which could be an influence in the spectra obtained in this study. As visual observations and observed higher values in the NIR have shown, a conclusion can be made that

later-released cultivars had more biomass than the older cultivars and that this additional biomass could be used for increased light interception (Figures 2.2 and 2.3).

The overall observed trend in reflection values is a decrease in the VIS and an increase in the NIR. These observations suggest a greater absorption of light incident upon the leaves by the major pigments associated with photosynthesis and a high refection of NIR light by the same pigments, as well as an overall healthier leaf and canopy structure. The significant correlation differences in the NIR between maturity groups are most likely due to morphological and phenological differences between cultivars. The MGIII experiment comprised cultivars that were more adapted to Kansas growing conditions. In the MGIV experiment, hot and dry conditions in 2011 and an early frost in 2012 did not allow certain cultivars to reach full maturity. In addition, earlier-released cultivars in the MGIV experiment had more lodging issues, which tended to confound long wavelengths in the NIR more than in the MGIII experiment.

Conclusions

As these observations indicate, significant differences in the VIS, red-edge, and NIR waveband regions were found among, both, MGIII and IV genotypes. The largest genotypic differences were observed in the red and red-edge portion so of the spectra in between both MG genotypes. The NIR was far less significant than the VIS, with only small portions between MG genotypes exhibiting significant genotypic differences.

Improvements in seed yield through cultivar development resulted in several changes in spectral curves. The VIS and red-edge regions of the spectrum had the most significant negative correlations with YOR. The newer released cultivars tended to have lower reflectance values in the VIS and red-edge spectra portions and higher values in the NIR portion of the spectra than older cultivars. Breeding has most affected spectral reflectance curves by reducing the VIS

portion of the spectra and extending the red-edge, resulting in a shift to lower reflectance values further into the NIR and then a sharp inflection to the NIR, with higher values in the NIR. In addition, a crossover occurs in new genotypes around 1150 nm, resulting in lower reflectance values in the transition from the NIR to the middle infrared. These results suggest that breeding advancement has had an impact on spectral reflectance curves and the areas that have been changed with breeding advancement may be exploited for further advancement. Caution should be applied to certain aspects of spectral data because weather and soil confounds may lead to inconsistent spectra. Nevertheless, spectral reflectance data may provide an indirect selection tool for increasing genetic gain in yield of soybean cultivars.

References

Adams, M.L., W.D. Philpot, and W.A. Norvell. 1999. Yellowness index: an application of spectral 2nd derivatives to estimate chlorosis of leaves in stressed vegetation. Int. J. of Remote Sensing 20:3663–3675.

Ashley, D.A., and H.R. Boerma. 1989. Canopy photosynthesis and its association with seed yield in advanced generations of a soybean cross. Crop Sci. 29:1042–1045.

Asrar, G., M. Fuchs, E.T. Kanemasu, and J.L. Hatfield. 1983. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. Agron. J. 76:300–306.

Austin, R.B. 1999. Yield of Wheat in the United Kingdom: Recent advances and prospects. Crop Sci. 39:1604–1610.

Babar, M., M. Reynolds, M. van Ginkel, A. Klatt, W. Raun, and M. Stone. 2006a. Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. Crop Sci. 46(2):578–588.

1132 Babar, M., M. Reynolds, M. van Ginkel, A. Klatt, W. Raun, and M. Stone. 2006b. Spectral
1133 reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and
1134 canopy temperature in wheat. Crop Sci. 46(3):1046–1057.
1135 Bull, C.R. 1991. Wavelength selection for near-infrared reflectance moisture meters. J. Agric.
1136 Eng. Res. 49:113–125.
1137 Buttery, B.R., R.I. Buzzell, and W.I. Findlay. 1981. Relationships among photosynthetic rate,
1138 bean yield and other characters in field grown cultivars of soybean. Can. J. Plant Sci.
1139 61:191–198.
1140 Calderini, D.F., M.F. Dreccer, and G.A. Slafer. 1995. Genetic improvement in wheat yield and
1141 associated traits. A re-examination of previous results and the latest trends. Plant Breed.
1142 114:108–112.
1143 Chang, J., D.E. Clay, K. Dalsted, S. Clay, and M. O'Neill. 2003. Corn (*Zea Mays* L.) yield
1144 prediction using multispectral and multidate reflectance. Agron. J. 95:1447–1453.
1145 Chappelle, E.W., M.S. Kim, and J.E. McMurtrey, III. 1992. Ratio analysis of reflectance spectra
1146 (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll a,
1147 chlorophyll b, and carotenoids in soybean leaves. Remote Sens. Environ. 39:239–247.
1148 Clevers, J.G.P.W. 1997. A simplified approach for yield prediction of sugar beet based on optical
1149 remote sensing data. Remote Sens. Environ. 61:221–228.
1150 Curran, P.J., J.L. Dungan, and H.L. Gholz. 1990. Exploring the relationship between reflectance
1151 red edge and chlorophyll content in slash pine. Tree Physiol. 7:33–48.
1152 Datt, B. 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total
1153 carotenoid content in eucalyptus leaves. Remote Sens. Environ. 66:111–121.

1154 Datt, B. 1999. Remote sensing of water content in Eucalyptus leaves. Australian Journal of
1155 Botany. 47:909–923.
1156 Daughtry, C.S.T., C.L. Walthall, M.S. Kim, E. Brown de Colstoun, and J.E. McMurtrey, III.
1157 2000. Estimating corn leaf chlorophyll content from leaf and canopy reflectance. Remote
1158 Sens. Environ. 74:229–239.
1159 Deering, D.W. 1978. Rangeland reflectance characteristics measured by aircraft and spacecraft
1160 sensors. Ph.D. diss., Texas A&M Univ., College Station.
1161 Dusek, D.A., R.D. Jackson, and J.T. Musick. 1985. Winter wheat vegetation indices calculated
1162 from combinations of seven spectral bands. Remote Sensing of Environment 18:255–267.
1163 Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa
1164 Agric. Home Econ. Exp. Stn., Iowa State Univ., Ames.
1165 Filella, I., I. Serrano, J. Serra, and J. Peñuelas. 1995. Evaluating wheat nitrogen status with
1166 canopy reflectance indices and discriminate analysis. Crop Sci. 35:1400–1405.
1167 Filella, I., and J. Peñuelas. 1994. The red edge position and shape as indicators of plant
1168 chlorophyll content, biomass, and hydric status. Int. J. of Remote Sensing. 15(7):1459-
1169 1470.
1170 Gamon, J.A., J. Peñuelas, and C.B. Field. 1992. A narrow-waveband spectral index that tracks
1171 diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41:35–44.
1172 Gamon, J.A., and J.S. Surfus. 1999. Assessing leaf pigment content and activity with a
1173 reflectometer. New Phytol. 143:105–117.
1174 Gao, B.C. 1996. NDWI—a normalized difference water index for remote sensing of vegetation
1175 liquid water from space. Sens. Environ. 58:257-266.

1176 Gitelson, A., and M.N. Merzlyak. 1996. Signature analysis of leaf reflectance spectra: Algorithm
1177 development for remote sensing of chlorophyll. J. Plant Physiol. 148:494–500.
1178 Gitelson, A.A., and M.N. Merzlyak. 1994. Quantitative estimation of chlorophyll-a using
1179 reflectance spectra: Experiments with autumn chestnut and maple leaves. J. Photochem.
1180 Photobiol. 22:247–252.
1181 Gutierrez, M., M. Reynolds, W. Raun, M. Stone, and A. Klatt. 2010. Spectral water indices for
1182 assessing yield in elite bread wheat genotypes under well-irrigated, water-stressed, and
1183 high-temperature conditions. Crop Sci 50:197–214.
1184 Hanson, P.M., J.R. Jorgenson, A. Thomsen. 2002. Predicting grain yield and protein content in
1185 winter wheat and spring barley using repeated canopy reflectance measurements and
1186 partial least squares regression. J. of Ag Sci. 139:307-318.
1187 Kumar, R., and L. Silva. 1973. Light ray tracing through a leaf cross-section. Appl. Optics
1188 12:2950–2954.
1189 Lin, W.S., C.M. Yang, and B.J. Kuo. 2012. Classifying cultivars of rice (*Oryza sativa* L.) based
1190 on corrected canopy reflectance spectra data using the orthogonal projections of latent
1191 structures (O-PLS) method. Chemometrics and Intelligent Laboratory Systems. 115:25–
1192 36.
1193 Ma, B.L., L.M. Dwyer, C. Costa, E.R. Cober, and M.J. Morrison. 2001. Early Prediction of
1194 Soybean yield from canopy reflectance measurements. Agron. J. 93:1227–1234.
1195 Marti, J., J. Bort, G. Slafer, and J. Araus. 2007. Can wheat yield be assessed by early
1196 measurements of normalized difference vegetation index? Annals of Applied Biology
1197 150:253–257.
1198 Moran, J.A., A.K. Mitchell, G. Goodmanson, and K.A. Stockburger. 2000. Differentiation

among effects of nitrogen fertilization treatments on conifer seedlings by foliar reflectance: A comparison of methods. Tree Physiol. 20:1113–1120.

Morrison, M.J., H.D. Voldeng, and E.R. Cober. 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. Agron. J. 91:685–689.

Moss, D.N., and R.B. Musgrave. 1971. Photosynthesis and crop production. Adv. Agron. 23:317–336.

Naes, T., T. Isaksson, T. Faern, and T. Davis. 2004. A user-friendly guide to multivariate calibration and classification. NIR Publ., Chichester, UK.

Peñuelas, J., F. Baret, and I. Filella. 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica. 31:221–230.

Peñuelas, J., I. Filella, C. Biel, L. Serrano, and R. Save. 1993. The reflectance at the 950– 970 mm region as an indicator of plant water status. International Journal of Remote Sensing. 14:1887–1905.

Peñuelas, J., R. Isla, I. Filella, and J.L. Araus. 1997. Visible and near infrared reflectance assessment of salinity effects on barley. Crop Sci. 35:1400–1405.

Prasad, B., B. Carver, M. Stone, M. Babar, W. Raun, and A. Klatt. 2007a. Genetic analysis of indirect selection for winter wheat grain yield using spectral reflectance indices. Crop Sci. 47:1416–1425.

Prasad, B., B. Carver, M. Stone, M. Babar, W. Raun, and A. Klatt. 2007b. Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions. Crop Sci. 47:1426–1440.

1221 Reynolds, M., Y. Manes, A. Izanloo, and P. Langridge. 2009. Phenotyping approaches for
1222 physiological breeding and gene discovery in wheat. Annals of Appl. Bio. 155(3):309–
1223 320.
1224 Reynolds, M.P., S. Rajaram, and K.D. Sayre. 1999. Physiological and genetic changes of
1225 irrigated wheat in the post-green revolution period and approaches for meeting projected
1226 global demand. Crop Sci. 39:1611–1621.
1227 Ries, L.L., L.C. Purcell, T.E. Carter Jr., J.T. Edwards, and C.A. King. 2012. Physiologial traits
1228 contributing to differential canopy wilting in soybean under drought. Crop Sci. 52:272–
1229 281.
1230 Royo, C., N. Aparicio, D. Villegas, J. Casadesus, P. Monneveux, and J. Araus. 2003. Usefulness
1231 of spectral reflectance indices as durum wheat yield predictors under contrasting
1232 Mediterranean conditions. Int. J. of Remote Sensing 24:4403–4419.
1233 Royo, C., D. Villegas, L.F. Garc´ıa del Moral,, S. El Hani, N. Aparicio, Y. Rharrabti, and J.L.
1234 Araus. 2002. Comparative performance of carbon isotope discrimination and canopy
1235 temperature depression as predictors of genotype differences in durum wheat yield in
1236 Spain.Aust. J. Agr. Res. 53:561–569.
1237 Sayre, K.D., S. Rajaram, and R.A. Fischer. 1997. Yield potential progress in short bread wheat in
1238 Northern Mexico. Crop Sci. 37:36–42.
1239 Serrano, L., I. Filella, and J. Peñuelas. 2000. Remote sensing of biomass and yield of winter
1240 wheat under different nitrogen supplies. Crop Sci. 40:723–731.
1241 Shanahan, J.F., J.S. Schepers, D.D. Francis, G.E. Varvel, W.W. Wilhelm, J.S. Tringe, M.R.
1242 Schlemmer, and D.J. Major. 2001. Use of remote sensing imagery to estimate corn grain
1243 yield. Agron. J. 93:583–589.

1244 Shibles, R. 1978. Adaptation of soybeans to different seasonal durations. In: R. J. Summerfield
1245 and A. H. Bunting, editors, Advances in Legume Science. Royal Botanic Gardens,
1246 University of Michigan. p. 279–285.
1247 Sims, D.A., and J.A. Gamon. 2002. Relationship between leaf pigment content and spectral
1248 reflectance across a wide range species, leaf structures and development stages. Remote
1249 Sens. Environ. 81:337–354.
1250 Sinclair, T. R., L.C. Purcell, and C.H. Sneller. 2004. Crop transformation and the challenge to
1251 increase yield potential. Trends Plant Sci. 9:70–75.
1252 Thomas, J.R., and H.W. Gausman. 1977. Leaf reflectance vs. leaf chlorophyll and carotenoid
1253 concentrations for eight crops. Agron. J. 69:799–802.
1254 Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation.
1255 Remote Sens. Environ. 8:127–150.
1256 Vina, A., and A. Gitelson. 2011. Sensitivity to foliar anthocyanin content of vegetation indices
1257 using green reflectance. IEEE Geoscience and Remote Sensing Letters 8:464–468.
1258 Waddington, S.R., J.K. Ransom, M. Osmanzai, and D.A. Saunders. 1986. Improvement in the
1259 yield potential of bread wheat adapted to Northwest Mexico. Crop Sci. 26:698–703.
1260 Weber, V.S., J.L. Araus, J.E. Cairns, C. Sanchez, A.E. Melchinger, and E. Orsini. 2012.
1261 Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants
1262 grown under different water regimes. Field Crops Research 128:82–90.
1263 Welles, J. M., and J.M. Norman. 1991. Instrument for indirect measurement of canopy
1264 architecture. Agron. J. 83:818–825.
1265 Wessman, C. A. 1990. Evaluation of canopy biochemistry. In: R. J. Hobbs and H. A. Mooney,
1266 editors, Remote Sensing of Biosphere Functioning. 135–156.

Wiegand, C., A. Richardson, D. Escobar, and A. Gerbermann. 1991. Vegetation indexes for crop assessment. Remote Sensing of Environment 35:105–119.

Table 2.1. Cultivars used for spectral data evaluation with year of release and two-year yield averages.

Cultivar	Year of Release	Yield	Cultivar	Year of Release	Yield
		t ha^{-1}			t ha^{-1}
	MGIII			MGIV	
AK (Harrow)	1928	2.67	Boone	1935	2.34
Calland	1968	2.95	Chief	1940	1.97
Dunfield	1923	2.32	Clark	1953	2.70
IA 3010	1998	4.30	Cutler	1968	2.74
Illini	1927	2.72	Douglas	1980	2.76
Lincoln	1943	2.51	Flyer	1988	3.48
MACON	1995	3.86	KS4694	1993	3.24
NE3001	2004	3.59	LD00-3309	2005	4.09
Private 3- 1	1978	3.80	Macoupin	1930	2.29
Private 3- 8	2002	4.08	Private 4- 1	1985	3.54
Private 3- 9	1989	4.03	Private 4- 4	2001	3.86
Private 3-12	1997	4.11	Private 4- 6	1980	3.56
Private 3-13	2004	4.44	Private 4-11	2000	4.04
Private 3-14	2007	4.63	Private 4-12	1973	2.92
Private 3-15	1983	2.99	Private 4-13	1984	3.60
Private 3-21	2001	4.30	Private 4-19	2006	3.68
Private 3-23	2006	4.49	Private 4-20	2008	4.21
Shelby	1958	2.81	Private 4-21	2010	4.26
Wayne	1964	2.92	Private 4-22	2000	3.68
Williams	1971	3.44	Sparks	1981	2.98

Table 2.2. Analysis of variance for seed yield of MGIII and IV experiments.

Source	D.F.	F Value	Pr > F
		MGIII	
Genotype (G)	18	111.57	< 0.01
Environment (E)	5	109.10	< 0.01
G x E	94	3.11	< 0.01
		MGIV	
Genotype (G)	19	95.98	< 0.01
Environment (E)	5	82.76	< 0.01
G x E	95	3.90	< 0.01

Table 2.3 Band regions with genotypic differences from analysis of variance (N=20), for MGIII and MGIV experiments.

	MGIII	MGIV
Band	F Value	F Value
405	20.61**	13.47**
415	19.96**	13.63**
425	17.41**	12.54**
435	15.80**	11.41**
445	15.26**	10.85**
455	15.11**	10.68**
465	15.41**	10.59**
475	15.78**	10.58**
485	16.09**	10.56**
495	16.90**	10.88**
505	18.66**	13.60**
515	22.73**	15.00**
525	29.10**	17.86**
535	32.83**	18.87**
545	34.42**	19.32**
555	34.77**	19.46**
565	33.83**	19.62**
575	31.08**	19.27**
585	28.34**	18.63**
595	26.62**	18.02**
605	25.23**	17.53**
615	22.78**	16.52**
625	21.12**	15.64**
635	20.46**	15.22**
645	19.29**	13.82**
655	17.94**	12.44**
665	17.11**	11.12**
675	17.33**	10.61**
685	17.52**	11.48**
695	22.96**	17.13**
705	37.14**	20.73**
715	38.07**	18.55**
725	22.66**	12.10**
735	8.69**	5.41**
745	3.12**	3.13**
755	1.96*	3.19**
765	1.85*	3.41**

775	1.96*	3.61**
785	1.99*	3.68**
795	2.00*	3.70**
805	2.01*	3.73**
815	1.97*	3.71**
825	2.00*	3.74**
835	2.04**	3.80**
845	2.10**	3.85**
855	2.12**	3.89**
865	2.15**	3.92**
875	2.17**	3.95**
885	2.16**	3.97**
895	1.97*	3.95**
905	1.59	3.96**
915	1.73*	3.96**
925	1.87*	4.01**
935	2.05*	3.84**
945	2.11**	3.93**
955	2.23**	4.01**
965	2.41**	4.14**
975	2.56**	4.23**
985	2.63**	4.30**
995	2.67**	4.34**
1005	2.67**	4.66**
1015	2.66**	4.68**
1025	2.65**	4.69**
1035	2.64**	4.70**
1045	2.64**	4.70**
1055	2.63**	4.70**
1065	2.61**	4.70**
1075	2.58**	4.70**
1085	2.34**	4.70**
1095	2.08**	4.69**
1105	2.39**	4.67**
1115	2.57**	4.61**
1125	2.60**	4.51**
1135	2.90**	4.65**
1145	3.38**	4.81**
1155	4.26**	5.08**
1165	4.56**	5.24**
1175	4.74**	5.32**
1185	4.86**	5.37**

1195	4.94**	5.41**
1205	4.95**	5.42**
1215	4.90**	5.41**
1225	4.84**	5.38**
1235	4.78**	5.37**
1245	4.73**	5.35**
1255	4.68**	5.34**
1265	4.65**	5.32**
1275	4.68**	5.34**
1285	4.75**	5.37**
1295	4.89**	5.38**
1305	5.00**	5.41**

* = $P \leq 0.05$

** = $P \leq 0.01$

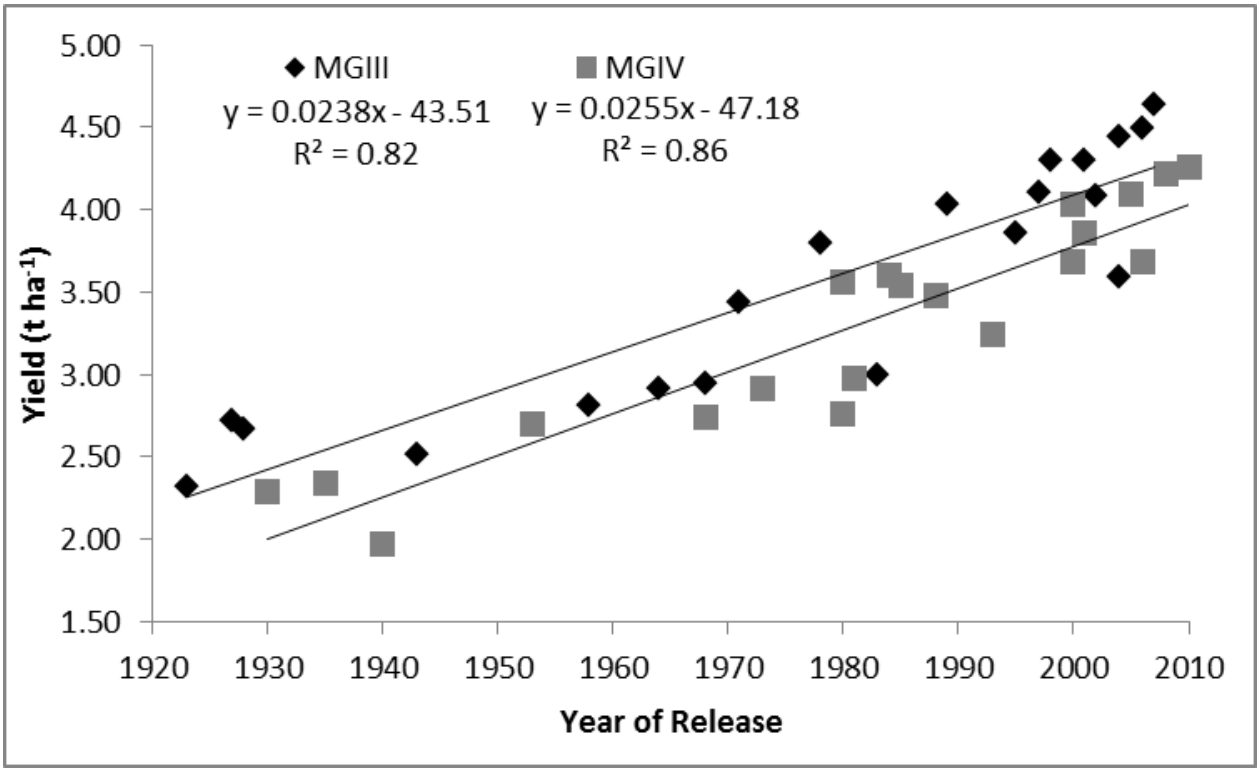


Figure 2.A. Relationship between year of release and two-year yield means for MGIII and MGIV experiments. Least square line with equation and coefficient of determination (R^2) represented at the top of the figure.

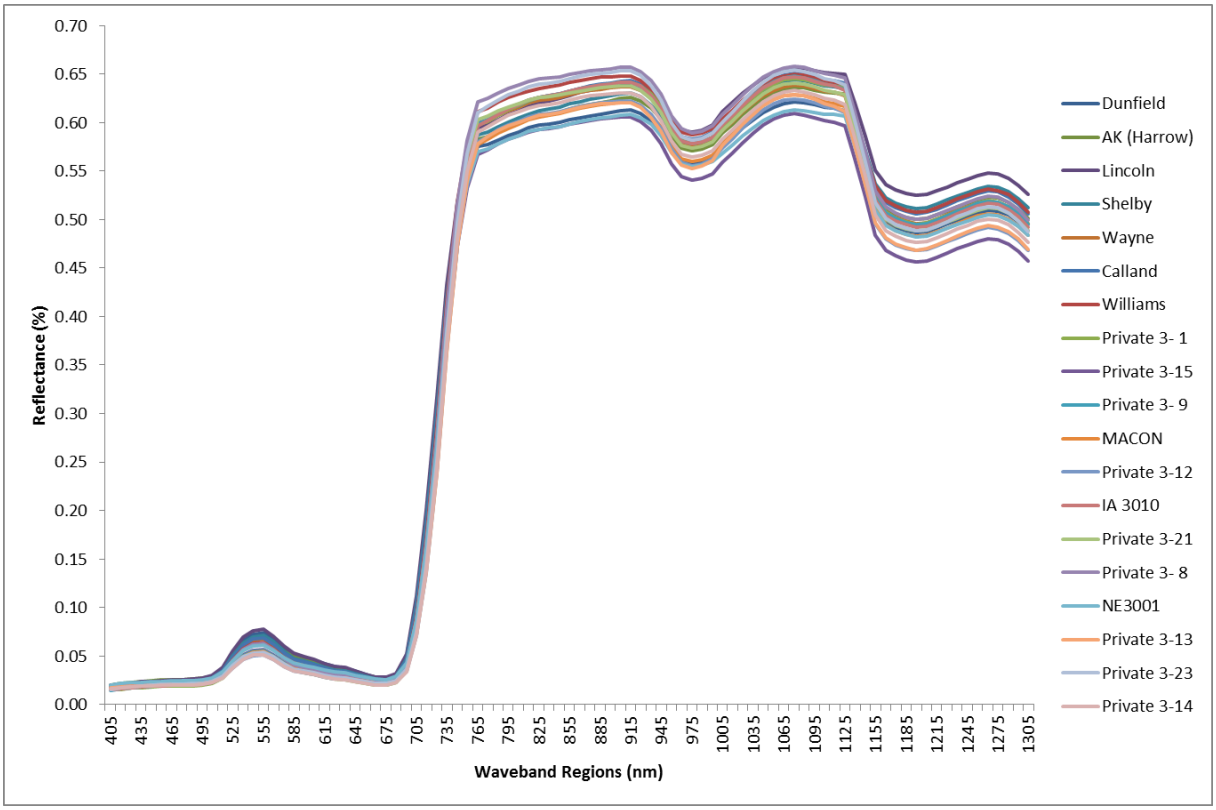


Figure 2.B. Mean spectral response curves of MGIII genotypes (without 'Illini'). Wavebands are 10nm intervals and reflectance is the percentage of a white reference panel.

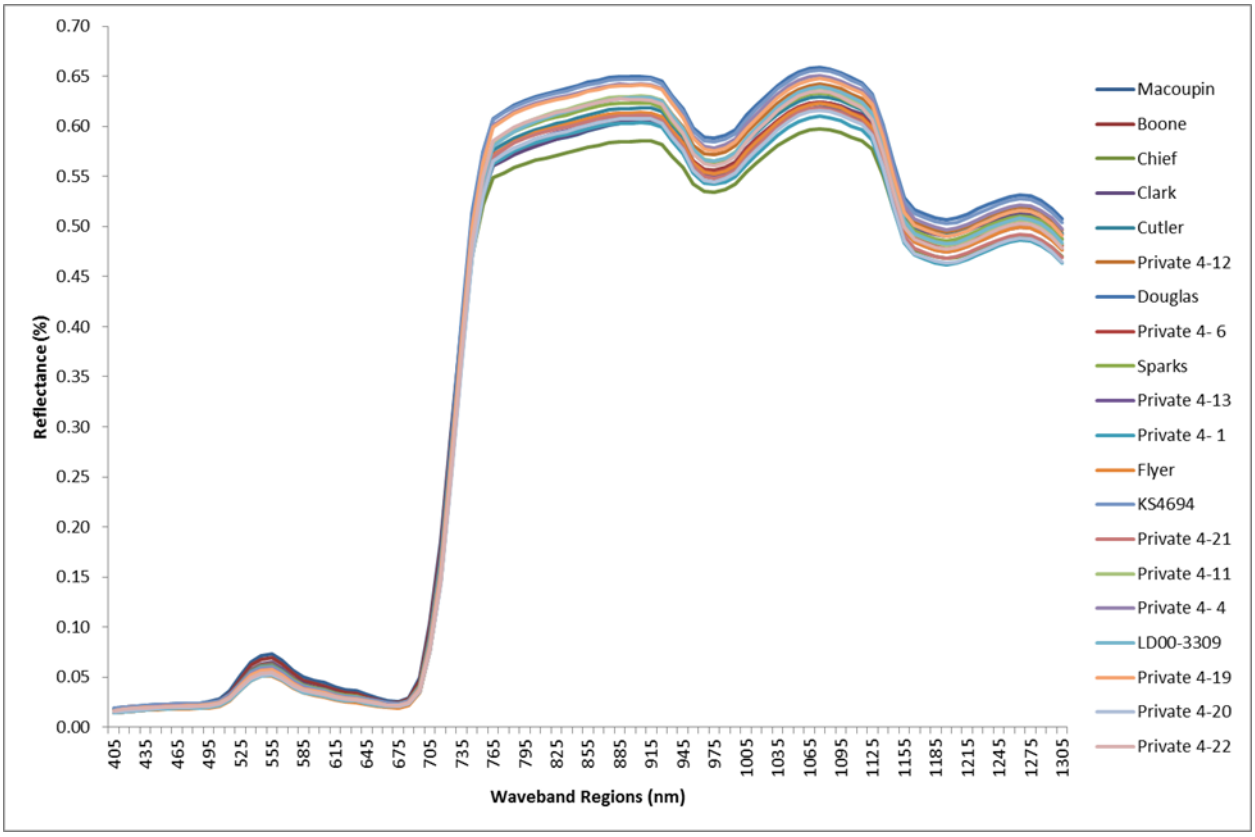


Figure 2.C. Mean spectral response curves of MGIV genotypes. Wavebands are 10nm intervals and reflectance is the percentage of a white reference panel.

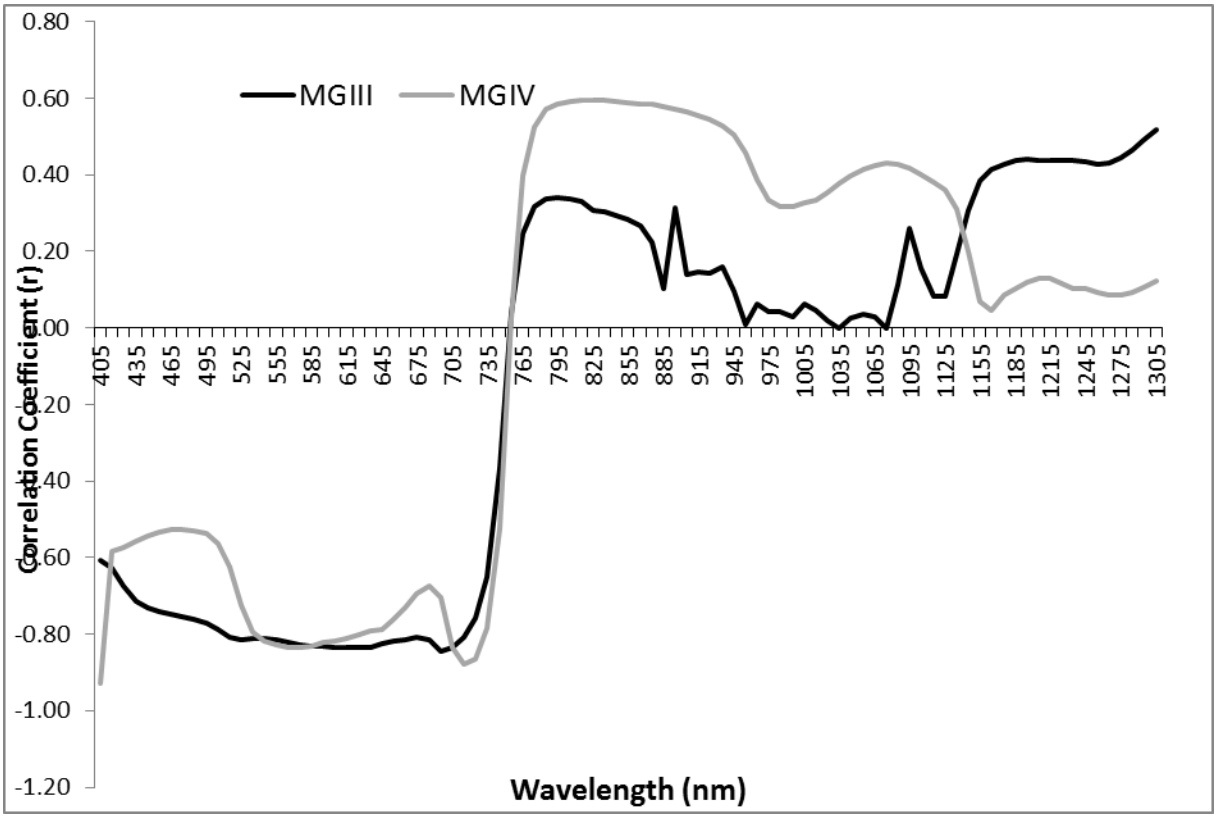


Figure 2.D. Correlation coefficient (r) values between wavelength (nm) and year of release for MGIII and MGIV two-year mean reflectance values. An r value ≥ ±0.37 is significant at the Pr≤0.05 level, and an r value ≥ ±0.56 is significant at the Pr≤0.01.

Chapter 3 - Characterizing Soybean Seed Yield Using Optimized Phenotyping with Canopy Reflectance

Abstract

Genotyping and phenotyping technologies that increase the amount of genetic material in the breeding program and increase the efficiency of cultivar development are necessary to accomplish the genetic gains needed to meet food demand. Optimized phenotyping using canopy reflectance measurements may provide a logical solution. The objectives of this study were: 1) to determine if canopy reflectance is useful in characterizing soybean seed yield; 2) to build yield estimation models for use as a screening tool; 3) to determine which wavebands contribute most to soybean yield estimation; and 4) to determine which spectral reflectance observations at specific growth stages contribute most to soybean seed yield. Canopy reflectance and seed yield were measured on 20 maturity group III (MGIII) and 20 maturity group IV (MGIV) soybean cultivars released from 1923 to 2010. Measurements were conducted on six irrigated and water-stressed environments in 2011 and 2012. Spectral band regions significantly contributing to seed yield prediction were selected through partial least squares regression, and yield estimation models were created using selected band regions through multiple linear regression. Significant differences were detected between genotypes, environments, and genotype × environment interactions for yield and band regions. Yield prediction models created using the red edge and portions of the near-infrared spectrum explained much of the variation in seed yield among genotypes. No significant trend was found for specific growth stages contributing more to yield estimation modeling or between water regimes. Yield estimation models using canopy reflectance measurements may be a useful selection tool in breeding programs.

Introduction

Due to the complex genetic basis of soybean yield, classic breeding approaches are still utilized to assess yield potential and improvement. In these approaches, crosses are made with elite lines and progeny and then selected based on attributes favorable to specific target environments. Superior genotypes are moved through the process, with seed yield used as a main selection criterion (Loss and Siddique, 1994). This selectivity has decreased genotypic diversity because many parental lines have similar backgrounds. Mechanized harvesters are utilized to make the process of obtaining seed yield faster, but this method is expensive, laborious, and time-consuming. Field evaluation of plots under normal growing conditions will always be necessary, but if specific regions of the electromagnetic spectrum can consistently account for a large portion of seed yield variability between genotypes and across environments, the technology may a useful indirect selection criterion for breeding programs.

Most of the field-based research conducted on yield estimation models using canopy reflectance are focused on one-, two-, or three-band indices, which can be highly variable and inconsistent (Babar et al., 2006a, 2006b). Some of these indices have been useful in estimating yield through other plant characteristics, such as chlorosis (Adams et al., 1999), green cover (Dusek et al., 1985; Daughtry et al., 2000), chlorophyll (Datt, 1999; Daughtry et al., 2000), and photosynthetically active tissue (biomass) (Wiegand et al., 1991). These indices have had varying degrees of success, but none have fully captured the underlying physiological and environmental factors behind consistent phenotypic yield estimation.

Studies focusing on utilizing full-spectrum instruments and models for various parameters have been published in recent years in wheat (Hansen et al., 2002; Pimstein et al., 2007, 2011), corn (Hong et al., 2001; Weber et al., 2012), rice (Lin et al., 2012), and cotton (Zhao et al., 2006) in optimal and drought environments. Additional studies have characterized

and explored how to incorporate these new models into high-efficiency platforms (Walter et al., 2012; White et al., 2012).

Most of the indices created correlate plant health parameters such as pigment status to grain yield. Indices such as the simple ratio (SR), first used by Jordan (1969) and Rouse et al. (1973) and described by the equation [R_{NIR}/R_{RED}]), capture the ratio of NIR reflectance to red reflectance. SR has been shown to correlate well with biomass, leaf area index (LAI), fractional photosynthetically active radiation (FPAR), ground cover, and wheat and soybean yield (Hatfield, 1983; Wiegand et al., 1991; Ball and Konzak, 1993; Price and Bausch, 1995; Ma et al., 2001; Royo et al., 2003; Hatfield and Prueger, 2010). The normalized difference vegetation index (NDVI) (derived by Deering (1978) and Tucker (1979) to estimate green biomass and intercepted PAR; defined as [R_{NIR}-R_{RED}/R_{NIR}+R_{RED}]) has been used to predict yield and other plant functions with many crops using hyperspectral and satellite imagery (Wiegand et al., 1991; Peñuelas et al., 1997; Ma et al., 2001; Shanahan et al., 2001; Royo et al., 2002; Royo et al., 2003; Prasad et al., 2007a, 2007b; Marti et al., 2007). Gittelson et al. (1996a) proposed the Green NDVI as a substitute for the high saturation point of the NDVI's red region. Shanahan et al. (2001) have successfully predicted yields in corn under normal growing conditions in Nebraska using the Green NDVI. Researchers found that normalizing the green and NIR relationship was highly correlated with grain yield, explaining 70 to 92% of yield variability at mid–grain fill (Shanahan et al., 2001). Ma et al. (2001) used the Green NDVI (R_{613}-R_{559}/R_{613}+R_{559}) to predict soybean yield and found a high correlation with soybean yields under irrigated conditions, which explained up to 80% of the variation within yield. The photochemical reflectance index (PRI), which is defined by the equation ($R_{550} - R_{531}$)/($R_{550} + R_{531}$), captures the normalized difference between the major green wavelength reflectance of the plant canopy and leaves, which can be

used to quantify radiation use efficiency (Gamon et al., 1992). Trotter et al. (2002) and Garbulsky (2011) used the PRI to assess nitrogen use efficiency and radiation use efficiency to distinguish genotypes that were superior to checks. These indices have been used estimate yield varying environments on many platforms.

Other indices focus on the biophysical parameters, such as water content and leaf area. Leaf area index (first illustrated by Tucker and Sellers, 1986) was developed to predict vegetation parameters such as green biomass and green leaf area (Babar et al., 2006a, 2006b). The LAI captures the genotypic differences between photosynthetically active radiation (PAR) either being absorbed by leaves with more area or being able to monitor important crop variables throughout the growing season (Clevers, 1997). Aparicio et al. (2002) have determined that LAI plays a large part in plant function and correlates well with yield prediction. Many water indices have been developed to relate reflectance measurements with success in the 850 and 970 nm regions with bread and durum wheat yield (Royo et al., 2003; Gutierrez et al., 2010). The water indices focus on the minor water absorption bands within the spectrum and capture relative water content based on absorption strength of water within the plant leaves. Capturing biophysical properties can mean discovering properties that influence yield production of field crops. These indices are easy to calculate and give researchers a way to easily handle the large amounts of data associated with such research; however, indices tend to be environment-specific and have a low correlation with yield outside of wheat (Ma et al., 2001; Gutierrez-Rodrigues et al., 2004; Prasad et al., 2007a; Gutierrez et al., 2010). Therefore, more complex models that use more of the spectrum need to be developed to capture the necessary variation within yield across different environments and crops (Pimstein et al., 2007 and 2011; Weber et al., 2012).

High-yielding cultivars tend to have high biophysical and biochemical properties that make distinguishing these genotypes using spectral indices unreliable (Baret and Guyot, 1991; Aparicio et al., 2002; Pimstein et al., 2007). Due to confounds such as biomass saturation, dense cover, high LAI, and high chlorophyll levels, full-spectrum models have been considered for yield estimation. These models have not been utilized, however, because of over-fitting and collinearity concerns from the large number of predictors and typically small sample size associated with spectral research (Pimstein et al., 2007). The main approaches used to meet the challenges of hyperspectral data analysis focus on reducing the correlation between predictor variables, causing multicollinearity and normalizing the variability within predictor variables from sample to sample. Researchers also use these new techniques to make the spectral response curves into meaningful spectral waveband regions.

Principal component analysis (PCA) has been used by researchers to build yield prediction models that explain 70 to 90% of the yield variability in maize (Hong et al., 2001; Chang et al., 2003) and 39% of the yield variability among soybean genotypes (Hong et al., 2001). Using six bands, regression explained almost 95% of the variability within maize yield. Artificial Neural Network Analysis has also been used in soybean to build prediction models that explain between 46 and 81% of the yield variability in soybean and 42 to 77% in corn (Kaul et al., 2005). Researchers also determined that reducing the dataset to 10-nm intervals reduced the noise within the spectra (Lin et al., 2012). Utilizing multivariate analysis is new to precision phenotyping research but has the potential to create yield prediction models and explain other important phenotypic parameters.

Partial Least Squares (PLS) was used for initial data pre-treatment and variable reduction before multiple linear regression techniques (Wold, 1966). Partial least squares was selected for

analysis due to the high redundancy in spectral data and subsequent multicollinearity between bands and “large p small n” phenomenon associated with spectral data analysis. Like PCA, PLS extracts successive linear predictors of y called latent variables or factors. The new variables aim to explain response variation and predictor variation; models are selected based on balancing these two goals. Hansen et al. (2002) compared three different PLS methods and concluded the N-PLS was the most consistent for yield and protein content estimation in wheat and barley, which explained up to 75% of the variation in protein content and up to 97% of the yield variation. Using different water regimes, Weber et al. (2012) similarly found that PLS explained a maximum of 40% of the variability within corn yield and prediction models explained more variability in water-limited environments than non-water–limited environments. Lin et al. (2012) used orthogonal projections to latent structure PLS in rice and distinguished three cultivars with 90% accuracy, but this result dropped to 80% when distinguishing one cultivar.

The goal of this study was to develop statistical models for soybean yield estimation using canopy reflectance measurements. The specific objectives were to assess canopy reflectance as a tool for soybean yield estimation, identify specific growth stages significantly contributing to yield estimation, and test the stability and utility of yield estimation models across varying environments.

Materials and Methods

Experimental and Field Design

A study was conducted on soybean [*Glycine max* (L) Merr.] in 2011 and 2012 at the Kansas State University Research farm (south of Manhattan, KS). In 2011, experiments (location A) were conducted on a well-drained silt loam soil of the Eudora and Belvue types. In 2012, location B was a well-drained silt loam soil of the Bismarckgrove-Kimo complex and location C was a well-drained silt-sandy loam soil of the Belvue and Eudora types.

Twenty maturity group III (MGIII) and 20 maturity group IV (MGIV) soybean cultivars, ranging in release year from 1923 to 2010, were selected for release year diversity out of the Soybean Genetic Gains Study, which contained 60 MGIII and 54 MGIV genotypes. Selected cultivars were a mixture of private and public released cultivars.

The experiment was planted in a randomized complete plot design with 4 replications in 2011 and 2012. Replicated plots were planted on 23 May, 2011 on field A, and 16 May, 2012 and on 4 June, 2012 on fields B and C respectively. Each experimental unit consisted of 4 rows, 3.4 meter long, and spaced 76 cm apart. Genotypes were planted in separate irrigated (IRR or I) and water stressed (DRY or D) environments. In both years, weed pressure and fertility were not limiting. In 2011 and 2012, flood irrigation was applied to the irrigated plots starting at reproductive stage 1 (R1), and continued weekly until R6 (Fehr and Caviness, 1977). In 2011, no supplemental irrigation was applied to the water-stressed environments whereas in 2012, due to the extremely dry conditions, one irrigation of approximately 5 cm was applied to the DRY plots on fields B and C shortly after R1 to ensure crop development. No further irrigation was applied to the DRY plots on field B. Weekly irrigation continued in the IRR plots. Later in the growing season on Field C, an irrigation malfunction occurred during seed fill, and the water-stressed

environment received an additional 5 to 6 cm of water. This irrigation occurred after all of the spectral readings had been taken. Because of this late-season irrigation, the water treatments for Field C are designated irrigation 1 (IRR1, or I1) and irrigation 2 (IRR2 or I2), rather than IRR and DRY.

Phenotypic Traits

Maturity, height, and lodging were taken on all plots during both seasons. Maturity was calculated as the number of days past 31 August when 95% of the pods had reached mature color. Height was measured as the average length (in cm) of mature plants from the ground to the tip of the main stem. Lodging was scored on a scale of 1 to 5 based on the number of plants leaning or broken. Upright plants with no lean were scored a 1, 2 = 20° lean, 3 = 45° lean, 4 = 60° lean, and 5 = flat on the ground. Plots were mechanically harvested using a two-row plot combine, with rows two and three used to calculate seed yield of the plot. Seed yield was calculated as the total mass of seed obtained and was adjusted to uniform moisture.

Data Analysis

Analysis of Variance

The GLIMMIX procedure in SAS 9.2 (SAS Institute, 2008) was used for analysis of variance (ANOVA). Cultivar and experiment were treated as fixed effects, and replication nested within the experiment was treated as a random effect for yield and agronomic traits. For hyperspectral data, replication nested within experiment and subsamples were treated as random effects for full-season analysis. For observation-day analysis, the repeated statement for repeated sampling data was used and treated as a random effect. Environment was classified as year, field, and water treatment. The Brown-Forsythe homogeny of variance test was implemented in all

ANOVA analyses, and residuals were used to determine outliers that were excluded from further statistical analysis.

Partial Least Squares (PLS)

PROC PLS in SAS 9.2 with the non-linear iterative partial least squares (N-PLS) algorithm, random k-fold = 10-cross validation, and factor levels set at 15 was used for outlier identification, pre-treatment, and predictor variable selection on 2011 and 2012 data. The predicted residual sums of squares (PRESS) were used for model selection. The NIPALS algorithm was chosen over the SIMPLS algorithm (De Jong, 1993) because of the high dimensionality and non-linear nature of the hyperspectral data. Random k-fold = 10-cross validation was chosen because of the small sample size compared with the number of predictors and the lack of a true validation dataset for validating built models. Max factor number was set at 15, which is the max factor number allowed by SAS for the dataset used. Optimized factor number was selected when the PRESS statistic was minimized most (Wold, 1966).

Outlier Identification and Data Synthesis

Hyperspectral data were initially trimmed from 350 to 2500 nm to 400 to 1310 nm to eliminate noise from atmospheric absorption regions focused around the water absorption bands in the near-infrared and atmospheric noise in the ultra-violet areas of the spectrum. For each observation day, initial outlier control on raw data was implemented before averages were calculated for observation day and season totals. After outliers were identified and excluded from analysis, data were combined to form 10-nm band regions to reduce the dataset size and eliminate some of the collinearity associated with bands in close proximity (Naes et al., 2004). It was also determined that combined bands would have less variation from sample to sample than single band measurements (Lin, et al., 2012).

Pre-treatment Before Modeling

Because modeling assumptions of normality were not met, it was necessary to auto-scale spectral data and dependent variable before PLS analysis (Harshman and Lundy, 1984). The data were mean-centered and auto-scaled separately for each individual experiment analysis and then across experiments when datasets were combined. Mean-centering and scaling eliminates bias within the bands to ensure even weighting of band magnitude, so the analysis is focused on the variance (Hansen et al., 2002). Auto-scaling is calculated as the inverse of the standard deviation for each variable [1/(SDev)].

Predictor Variable Selection

Hyperspectral band regions from 2-yr means and individual environments that most significantly contributed to yield estimation were determined using partial least squares regression (PLSR). In total, 91 band regions from 400 to 1310 nm were analyzed for variable importance and selected for further analysis through traditional least squares multiple linear regression (MLR). This procedure was done to extract exact band regions contributing to yield estimation, which is not possible in PLS models; therefore, custom sensors based on specific bands could be built for further research studies. Variables were determined to be important through variable importance within projection plots (VIP), which were created through a macro in SAS. Wold's criterion limit of 0.8 was used as a cutoff point to indicate variables that contributed significantly to yield estimation (Wold, 1966). Band regions were selected within peaks and over Wold's threshold. Due to collinearity issues, band regions that significantly contributed to yield estimation and were close to each other were combined for final analysis.

Multiple Linear Regression

PROC GLMSELECT in SAS 9.2 was used for yield estimation model building. All models were built using band regions selected through PLSR. For variable selection, stepwise with forward and backward elimination was used with a 0.1 alpha for variable entry and exit criterion. Due to the small sample sizes, full n-1 cross-validation (jack-knife) was utilized instead of training and validation datasets. Model selection was completed when the predicted residual sums of squares (PRESS) statistic and Mallow's Cp statistic were fully minimized. The selected beta parameter estimates were then exported into PROC REG (SAS 9.2) for goodness of fit tests. The variable inflation factor (VIF) was used to determine multicollinearity between parameters, and a selection value of greater than 10 was used to determine multicollinearity. If a parameter had a VIF greater than 10, it was eliminated from the possible variables, and the models were re-optimized.

All models were created using total-season least squares mean estimates with maturity groups kept separate. Two datasets (consisting of least square means from observation days determined significant contributors to yield estimation) were created for model selection. Each observation day was individually regressed with yield, and significant days were selected based on R^2 values, rMSE, and percentage of the dependent mean accounted for by the rMSE. Models were validated on individual environments and maturity group for reliability and robustness.

Results and Discussion

Genotypic Performance in MGIII and MGIV Yield and Spectra

Genotypes and genotypic performance for yield are the same as in chapter 2. A wide range was observed in both maturity groups for mean seed yield. Average seed yield ranged from 2.32 to 4.63 t ha^{-1} and 1.97 to 4.26 t ha^{-1} across environments for the MGIII and MGIV

genotypes, respectively. Mean yields and yield ranges and differences between water regimes varied highly between locations (Table 3.1). For yield, highly significant genotypic, environment, and genotype × environment (G × E) interaction differences were detected in both maturity groups (Table 3.2). In fields A and B, where the water regimes were maintained throughout the growing season, seed yields were significantly higher in the irrigated treatments than in the water-stressed treatments. In Field C, where the breakdown of the water regimes occurred, the irrigated and water-stressed yields were not statistically different.

Analysis of 10-nm hyperspectral band regions revealed genotype and environmental differences among many waveband regions. The regions that were significant for genotype, environment, or G × E are listed in Table 3.2. Maturity exhibited significant genotypic, environment, and G × E interaction differences in both maturity groups (Table 3.2). The G × E interaction did not account for a large portion of the total phenotypic variation, which suggests that genotypic means observed in a single environment could be replicated across environments with minimal interaction influence. Looking at individual growth stage observations, the visible VIR and red-edge bands were more consistent across growth stages and environments for genotypic differences in both maturity groups (Table 3.3). In the MGIII and MGIV genotypes, wavebands at 415 to 715 nm exhibited highly significant ($Pr > 0.01$) and significant ($Pr > 0.05$) genotypic differences in all environments and growth stages; however, no wavebands in the NIR had significant genotypic differences in all environments and growth stages in both maturity groups.

For the MGIII genotypes, the later growth stages are more consistent than earlier growth stages for genotypic differences for all wavebands; however, in the MGIV genotypes, no such trend is observed, with early and late growth stages exhibiting genotypic differences. As seen,

the G × E in the VIR and red-edge wavebands is minimal, whereas in the NIR, the G × E is greater in both maturity groups. Lin et al. (2012) observed greater genotypic differences in the NIR compared with the VIR in rice. This could be due to the high biomass variability in their study, which mainly affects the NIR. The small contribution that the G × E interaction contributes to the total phenotypic variation suggests that spectral data may be robust across environments, making it a strong candidate for an indirect selection technique. Reynolds et al. (2009) suggested that G × E stability is one of the major factors for evaluating any new phenotyping technique.

Waveband Region Selection

Using PLSR portions of the 2011 and 2012 mean spectra significantly contributed to yield estimation within the MGIII and MGIV experiments (Figures 3.1A and 3.1H). The VIR (400 to 700 nm) and red edge (701 to 730 nm) contributed significantly to yield estimation, with the red and red-edge portions of the spectra exhibiting the most significant projection to yield estimation in both maturity groups. Most of the NIR portion of the spectra did not contribute significantly to yield estimation for either maturity group; however, in both maturity groups, the 1290- to 1310-nm regions of the spectra were slightly above the 0.8 threshold, which suggests increased significance for yield estimation in these regions.

For 2011 Field A irrigated and water-stressed environments, both MGIII and MGIV mean spectra exhibited high variable importance values in the VIR and red-edge portions of the spectra (Figures 3.1B, 3.1C, 3.1I, and 3.1J. For the MGIII environments (A-I), visible and red edge portions had the highest significances for yield estimation, with a sharp decrease at the beginning of the NIR and a significant increase from 775 to 955 nm (Figure 3.1b). For the A-D environment, variable importance was similar to that of the A-I environment, but a decrease in

the blue (400 to 500 nm) portion of the spectra was observed, and there were significant peaks from 1035 to 1095 nm (Figure 3.1C).

For Field B, the VIR and red edge were highly significant in 2012, similar to that of Field A and the 2-yr averages for both maturity groups. For environment MGIII B-I, the VIR and NIR had highly significant portions of the spectra for yield estimation (Figure 3.1D). The red edge was the most significant, followed by the green and red. A spike in the NIR was observed around the 915-, 1100-, and 1165-nm regions, as was a sharp peak in the early blue portion of the spectra. For the MGIII B-D environment the VIR and NIR had highly significant regions above Wold's 0.8 threshold, with a sharp decrease in the NIR portion until 1165 to 1305 nm. The red edge portion was the most significant, as in the A–I environment (Figure 3.1E).

In 2012, neither maturity group in Field C was as consistent as in the other two fields; there was an irrigation malfunction, no spectral data was collected on IRR1 after the irrigation malfunction, and yield values were higher than expected. In addition, after the irrigation malfunction, plants were highly lodged, which could have contributed to inconsistent spectra influenced by soil background and sensing of under-leaf or stem reflectance. As in the other fields and environments, however, the VIR and red edge portions were the most useful for yield estimation (Figures 3.1F, 3.1G, and 3.1M).

The MGIII C-I environment had high peaks of significance in the VIR, but these peaks were not as strong as those in locations A and B (Figure 3.1F). There was a sharp increase in the red edge and slight peaks in the 915- to 1005-nm region, as well as in the 1165- to 1305-nm regions of the NIR. The C-I 2 environment mostly mirrors the C-I 1 environment, with a sharp peak in the green in the 550-nm region (Figure 3.1G). Compared with the C-I 1 environment, the

C-I 2 environment saw a sharp increase in the early blue, and the red edge was the most significant for yield estimation (Figure 3.1G).

In the MGIV A-I environment, a low significance in the blue region and a sharp increase in the green region was observed, with 550 nm being the highest significance for yield estimation in the green (Figure 3.1I). The red portion of the spectra was not as significant as the green or as in the MGIII environments; overall, however, the red-edge was the most significant region for yield estimation, similar to the MGIII environments. There was also a sharp decrease in the beginning of the NIR, with the rest of the plateau region (735 to 910 nm) being significant. The MGIV A-D environment had an increase in the blue region compared to the A-I environment, with the green being the most significant region of the VIR, similar to the A-I environment (Figure 3.1J). The red edge was significant, as in the MGIII and MGIV A-I environment; however, no significant regions of the NIR were observed.

In the MGIV B-I and B-D environments, significant portions of the spectra in the blue, green, red, and red edge were observed, as well as some portions of the NIR (Figures 3.1K,3.1 L). For the B-I environment, the early blue, green, red, and red edge were the most highly significant for yield estimation, as well as a slight peak in the 1305-nm region (Figure 3.1K). For the B-D environment, the early blue, green, and red edge were the most highly significant, similar to the B-I environment (Figure 3.1L); however, the red portion of the spectra and the NIR had different significant levels than the B-I environment. In the NIR, peaks in the 765 to 945 nm and a slight peak from 1065 to 1105 nm were observed, which were not observed in the B-I environment. These regions have been correlated with biomass production was well as plant water content; this correlation suggests that genotypes in the water-stressed environment may be differentiating in total biomass and water content, which leads to differences in yield.

For the MGIV environments in Field C, no optimized model was found for the C-I 2 environment because of higher than expected yield and inconsistent spectra from irrigation issues. The C-I 1 environment, however, had significant regions in the VIR and red edge, as well as most of NIR (with the exception of 1005 to 1125 nm) (Figure 3.1M). Overall, for the MGIV experiments, Field C was the most inconsistent and had the most variability between regions significant for yield estimation.

Different wavebands between maturity group mean spectra were detected for yield estimation. The differences in spectral regions correlated to yield could be due to morphological characteristics of the maturity groups. Other spectral regions exhibited variable importance values over Wold's criterion but were not selected due to fears of high multicollinearity between bands and over-fitting due to more predictor variables than the sample size (N = 20). Overall, the most significant portions of the spectra across all experiments were the visible and red-edge portions of the spectra. The green around the 550-nm region and red around the 675–695-nm regions were the most significant portions of the visible spectra. In the red edge, the 705–715-nm regions were the most significant. The NIR was highly influenced by environmental factors and was inconsistent in significance for yield estimation across environments. This result is most likely due to atmospheric scatter and observational day conditions influencing the spectra. Band regions for final yield modeling were selected based on significance to yield estimation through importance in projection values in all environments in both maturity groups. Selected portions of the spectra that were close to each other were combined to form 11 spectral regions used for further yield modeling. The final bands used for modeling were 415 (400–430 nm), 550 (530–570 nm), 680 (670–690nm), 715 (700–730 nm), 915 (910–920 nm), 940 (930–950 nm), 990 (980–1000 nm), 1100 (1090–1110 nm), 1140 (1120–

1160 nm), 1245 (1240–1250 nm), and 1300 (1290–1310 nm). The 405–435-nm regions have been correlated with high chlorophyll a and b as well as beta-carotene absorption, resulting in low reflectance values in soybean (Chappelle et al., 1992). Lower reflection values in this region are due to high absorption of light by the chlorophylls and beta-carotene. Peñuelas et al. (1995) also created a normalized phaeophytinization index (NPQI) that related senescence to reflectance in the 415- and 435-nm regions.

The 535–565-nm regions of the spectra are in the visible green region of the spectra, which has high reflection due to chlorophyll a and b reflecting green light. Chappelle et al. (1992) found that chlorophyll a and b of soybean leaves reflected the 550-nm region the most and could be used in a ratio with 675 nm to explain 93% of the variation within chlorophyll content between soybean leaves and could then be related directly to photosynthetic capacity. Also, 550 nm has been used to explain 92% of yield variability in wheat yield (Royo et al., 2003). Ma et al. (2001) also used 559 nm by itself as well as in a ratio with 613 nm to explain from 13 to 80% of the variation within yield in soybean genotypes.

The 675–685-nm regions of the spectra are in the middle of the red depression region of the spectra and have been correlated with chlorophyll absorption in soybean (Chappelle et al., 1992). The lower yielding varieties had a higher reflection value in this region, suggesting a lower amount of chlorophyll or inefficiency of the chlorophyll, resulting in lower yields in soybean (Ma et al., 2001), corn (Weber et al., 2012), and wheat (Royo et al., 2003).

The 705-745nm region encompasses the end of the red and start of the red-edge inflection point. The red-edge inflection point is the sharp increase in reflection values, due to the transition from the visible red region to the high-reflection NIR portion due to cellular

scatter. This inflection point has been used to distinguish plants for health and yield, with healthier plants having a large contrast between red and NIR (Gitelson et al., 2011).

The 915-nm waveband regions had a high reflection value in the spectra and have been associated with chlorophyll content measurements. Zhao et al. (2006) found that ratios using 551 and 915 nm as well as 708 and 915 nm accounted for 67–76 % of the variability within chlorophyll content between cotton genotypes; however, Gitelson et al. (2003 and 2005) found that the best wavebands for estimating chlorophyll in higher plants were from 525 to 585 nm and 695 to 725 nm. 915 nm also has been used for total biomass prediction in bermudagrass, accounting for 29.8 to 44.3% of the biomass variation (Stark et al., 2006). Marti et al. (2007) found that total biomass had a strong correlation with wheat yield ($r = 0.97$).

Reflection in the 940-nm region has been used in chlorophyll meters (Minolta Osaka Co., Ltd., Japan) to capture nitrogen status and chlorophyll content of crops (Blackmer et al., 1994). Vollmann et al. (2011) found a significant correlation with SPAD-502 readings (ratio between 650 nm and 940 nm) and 1000-seed weight in soybean cultivars. High reflection values were also observed in the 985–995-, 1105-, and 1135–1155-nm regions, with higher yielding genotypes tending to have higher reflection (data not shown). Wenjiang et al. (2004) found that wavebands selected through regression techniques for winter wheat total foliar nitrogen content were in the 1000–1140-nm ($r = 0.8325$) and 1200–1300-nm ($r = 0.5138$) regions.

The 1240-nm region of the spectra has been correlated with water content of the leaf in many crops (Peñuelas et al., 1993; Gao, 1996; Datt et al., 2003; Gutierrez et al., 2010). Moreover, lower values in the 1150–1260-nm region have been associated with higher water content (Sims and Gamon, 2003); Prasad et al. (2007b) concluded that indices using water content bands had higher heritability and could distinguish higher yielding genotypes more

consistently than vegetation based indices in wheat. The 1095-nm and 1295–1300-nm regions of the spectra are in the NIR portion, and higher values in these regions correlate to higher yielding varieties in corn, but no physiological characteristics have been associated with these waveband regions (Weber et al., 2012).

Correlation Between Parameters

Significant correlations were found between yield, maturity, and wavebands (Table 3.4). There was a significant positive correlation (r = 0.68) between yield and maturity in the MGIV experiment, but no significant correlation (r = 0.35) was found between yield and maturity in the MGIII genotypes. This result may have been mainly due to premature death of some MGIV entries in 2011 and an early frost in 2012 that did not allow these genotypes to develop fully. When the top five maturing cultivars were eliminated from analysis, no significant correlation was observed between yield and maturity in the MGIV genotypes (data not shown). In both maturity groups, the VIR wavebands had a negative and more significant correlation with yield than the NIR wavebands with the exception of 915 in the MGIV (r = 0.46*). The NIR portion had a positive correlation with yield, with the exception of 1245 and 1300 in both experiments. Chang et al. (2003) found similar patterns in correlation between VIR and NIR bands to corn yield; however, they found that in early sampling dates, the NIR had a negative correlation with yield. They concluded that negative correlations were due to soil reflectance confounds, similar to observations by Ryerson and Curran (1997). For the MGIII entries, 715 (r = -0.83**) had the most significant correlation with yield, and 1140 (r = 0.06) exhibited the least association (Table 3.4). In the MGIV experiment, 680 (r = -0.78**) was the most significantly correlated with yield, and 1140 (r = 0.03) was the least correlated.

Significant correlation coefficients were identified between maturity and the wavebands in both maturity groups (Table 3.4). For the MGIII genotypes, no significant correlations were identified in the VIR, but significant positive correlations were identified in the NIR (r = 0.51* to 0.73**). For the MGIV genotypes, significant correlations between maturity and wavebands were found in the VIR (r = -0.48* to -0. 59**) and NIR (r = 0.46* to 0.57**) portions of the spectrum. Due to the NIR's association with cell structure, earlier maturing varieties would have cellular degradation due to senescence sooner than later maturing varieties and would be distinguished due to low NIR values.

In both maturity groups, wavebands in the VIR portion were significantly correlated with each other (r = 0.77** to 0.99**), and the NIR wavebands were significantly correlated with each other (r = 0.69** to 1.00**). In the MGIII and MGIV genotypes, wavebands close in proximity were highly correlated with each other, and 1245 and 1300 had a correlation coefficient of 1.00. In both maturity groups, 1245 and 1300 tended to be significantly correlated with other bands, suggesting multicollinearity between these bands that could confound regression analysis. The high correlation between wavebands suggests that these wavebands are not independent information and techniques need to be employed to detect multicollinearity in regression analysis to reduce the risk of over-fitting yield estimation models.

Growth Stage Selection

Yield estimation models based on growth stages that explained a significant proportion of the yield variation were created for both maturity groups (Tables 3.5 and 3.6; growth stage observations selected for further analysis are bolded). Overall, some observation days were significantly better than other observation days. Weather and other environmental conditions seemed to affect the spectral response curves from observation day to observation day, leading to

some observations being inconsistent and showing lack of fit, whereas others significantly contribute to yield estimation. Also, expanding upon the intensity of observations to alleviate some of these problems may be necessary. The results seem to suggest that singular observation days are not consistent enough for yield estimation across different environments and recommendations for best growth stages for yield estimation cannot be concluded. Although not recorded in this study, these inconsistencies across days and environments could be due to changing air temperature, sun angle, and humidity levels affecting spectral reflectance values. Environments A-I and B-D tended to have the most consistent models, with all observation days in both maturity groups selected for final model building. Irrigated environment models tended to explain a greater portion of yield among genotypes than water-stressed environments, which is most likely due to genotypes being easier to distinguish in high-yielding environments than low–yield potential environments. No real trend for early or later growth stages was observed in this study, which is inconsistent with observations made by Ma et al. (2001), who found that late seed fill (R6) accounted for the most yield variability in soybean cultivars and that reproductive stages were better for yield estimation than vegetation stages. A study by Ma et al. (2001), however, was conducted using the Green NDVI, which may be more consistent in later growth stages than earlier growth stages due to greater genotypic differences in the green and NIR portions of the spectrum later in the growing season. Similar wavebands were selected for yield estimation in both maturity groups, with 715 (11-MGIII; 12-MGIV) and 915 (5-MGIII; 5-MGIV) utilized the most and 1300 (1) and 1245 (1) the least.

MGIII regression models for yield estimation based on growth stage explained from 47 to 85% of the variability in seed yield using 1–4 wavebands, with the exception of three environments with no significant regression models identified (Table 3.5). Root means square

error values ranged from 0.32 to 0.56 t ha^{-1}. The R^2 for each water regime explained a similar amount of yield variability in irrigated environments (R^2 = 0.47 - 0.85) and water-stressed environments (R^2 = 0.51 to 0.79) when comparing growth stages. Environments in 2012 (R^2 0.47 to 0.85) tended to explain more of the variation in yield than 2011 (R^2 = 0.51 to 0.73) environments; however, environment A-I had all observation days contributing to the final training model, whereas no other environment besides B-D had more than two observations contributing to the final training model dataset. For Field A, the R3-R4 growth stage in the AWI environment exhibited the highest R^2 for seed yield. The highest yield variability explained was in environment C-I 1 at the R6 (late seed fill) growth stage (R^2 = 0.85). The R3 and R3-4 growth stage observation on B-I and R3 growth stage observation on C-I 1 were not significant.

MGIV regression models for yield estimation based on growth stage explained from 30 to 89% of the variability in seed yield using 1–5 wavebands, with the exception of three environments with significant regression models identified (Table 3.6). Root mean square values ranged from 0.31 to 0.66 t ha^{-1}, with the R5 growth stage in the A-D environment exhibiting the lowest and the R4 growth stage in B-I exhibiting the highest values. Unlike the MGIII experiment, 2011 environments (R^2 = 0.54 to 0.89) tended to account for a larger amount of yield variability than 2012 data (R^2 = 0.30 to 0.86), and environment A-I was also the most consistent, with all observation days selected for final training model creation. The regression model created using the R3-R4 growth stage on A-I exhibited the highest coefficient of determination value, but also used the most wavebands (5), suggesting the model is most likely over-fit and the R^2 value was inflated. The same conclusion could be made about the R5 observation on B-I, but these growth stages were kept due to similar wavebands used in other models. The lack of fit observed for the A-I 2 environment can be explained by the previously explained irrigation malfunction. Also, as

in the MGIII experiment, the R2-R3 observation for B-I was discarded due to inconsistent spectra, which may have been related to high humidity and temperature conditions during spectral measurements.

Yield Estimation Model Development

Selected observation days from the individual environment analysis were combined into a single dataset for each maturity group and used to build the final training models for yield estimation in both maturity groups. Datasets were means calculated from singular observation day data. The summary statistics and equation for the selected growth stage observations based on two-year spectra and yield means for MGIII and MGIV genotypes are presented in Table 3.7 and Figure 3.2. For the MGIII genotypes, through multiple linear regression, the estimation model explained 83% of the variability within yield between genotypes. Red-edge region 715 explained the most yield variability in models (73%), and the 1100 waveband region explained 10% of the yield variability. In the training model, the β parameter estimate for the 715 waveband was negative, suggesting varieties with high values for the 715 will have decreased yields (Figure 3.2). This result is consistent with previous research findings that lower reflection values in the red region of the spectra correlates with higher grain yields (Weber et al., 2012). The 1100 waveband was positive, which is consistent with previous research suggesting increased reflectance in the NIR area can correlate with higher yields (Weber et al., 2012).

For the MGIV experiment, the training model accounted for 81% of the yield variability between genotypes. The 715 (700–730 nm) and 915 (910–920 nm) wavebands were selected in the training model, which is consistent with the red edge and NIR measurements seen in the MGIII training model. 715 explained the most yield variability at 70%, and 915 wavebands explained 11% of the variability in yield. The β parameter estimate for the 715 waveband was

negative, suggesting genotypes with high values for the 715 will have decreased yields, as in the MGIII models. The 915 waveband was also positive, which was consistent with MGIII genotypes and previous research (Reynolds et al., 1999; Lin et al., 2012; Weber et al., 2012).

Results suggest that selecting significant growth stage observations improved yield estimation models. Results also indicate that the 715 waveband explains a large portion of the variability within seed yield, and previous research concluded that soybean yield can be directly related to chlorophyll content (Morrison et al., 1999). Although not as important as chlorophyll, this study found wavebands in the NIR to be highly effective in estimating yield when combined with the red edge. Observations by Morrison et al. (1999) and Voldeng et al. (1997) that increases observed in soybean seed yield can be attributed to increased chlorophyll content and photosynthetic capacity. Reynolds et al. (1999) also concluded that increases observed in wheat yield were due to a better partitioning of photosynthetic components; however, Lin et al. (2012) concluded that the NIR regions from 760–1030 nm contributed the most to rice cultivar discernment.

Yield Estimation Model Validation

Validation of training models on individual environments delivered mixed results for both maturity groups, with coefficient of determination values ranging from 29 to 79% for the MGIII genotypes and 1 to 83% for the MGIV genotypes (Table 3.8). As in the observation-day selection analysis, the training model accounted for a large portion of the variability in yield, whereas in other environments and observation days, the training models did not account for a significant portion of the variability in yield. For the MGIII genotypes, the mean R^2 for the irrigated water regime was 56%, whereas in the water-stressed regime, the mean was 63%. Similar results were observed in the MGIV genotypes, with the irrigated regime accounting for

49% of the variability in with the C-I 2 environments and 59% without. The water-stressed environment mean accounted for 62% of variability, and when comparing water regimes within the same field, the water-stressed environment mean accounted for a larger portion in three of the four instances, with A-I for the MGIV genotypes as the lone exception. These results are similar to observations by Ma et al. (2001) that stressed environments accounted for more variability than optimal environments. The authors concluded that results were due to greater differences in genotypes between higher and lower yielding spectra in the non-optimal compared with the optimal environments. In three of six environments for MGIII and four of six environments for MGIV genotypes, the season totals (ST) accounted for as much or more of the variation in yield than individual growth stages. These results are somewhat surprising because season total means would be expected to be more robust and have less error associated with the values than single observations, resulting in higher R^2 values.

For the MGIII genotypes, the highest R^2 value was observed in the C-I 1 environment (R^2 = 0.79) season total dataset. The growth stage observation with the highest validation R^2 was observed in the C-I 1 environments as well, with an R^2 value of 0.75. The year averages were consistent for the MGII genotypes, with each year spectra mean dataset accounting for 68% of the variability within yield between genotypes. For the MGIV genotypes, the highest R^2 observed was the season total validation dataset for the B-D environment (R^2 = 0.83). The highest single growth stage was observed for the R1–R2 growth stage for environment B-I (R^2 = 0.79). Contrary to the MGIII training model, the MGIV training model did not have consistent performance for year means, with the 2011 mean dataset accounting for 65% and the 2012 mean dataset accounting for 79% of the variability within yield. This is surprising given that single growth stage validations on Field A in 2011 were more consistent than 2012 validations;

however, a larger difference was observed in validation R^2 values between water regimes for Field A than in Field B. Field C, of course, had the largest differences due to previously explained irrigation issues and the failure of spectral data to be representative of the harvested plants.

Regarding relative performance, models ranked the genotypes in a similar fashion in both maturity group experiments. Within a breeding program, rank order is usually more important than actual yield when comparing genotypes to commercial checks. This is due to the high G × E interaction and low heritability associated with yield. For the MGIV environment season totals, the top three highest yielding genotypes were distinguished in the top 25% (5 genotypes) in environments B-I and A-D as well as two-year means of the irrigated and water-stressed environments. This means that five genotypes would have to be selected select the top three yields. The top three genotypes were distinguished in the top 50% (10 genotypes) in environments B-D, C-I 2, and A-I. For the MGIV environments, the top three yielding genotypes were distinguished in the top 25% in the B-I and B-D environments and in the top 50% in the C-I 1, B-D, A-I, and two-year irrigated mean. Environment C-I 2 and the two-year water-stressed mean did not distinguish the genotypes within the top 75% of the genotypes. This is most likely due to problems stated above. In 14 of the 16 environments, the top three highest yielding genotypes were distinguished in the top 50% of the genotypes, and six of those were within the top 25%.

Conclusions

PLS variable selection was used to detect important spectral waveband regions contributing to soybean yield and reduced the spectral datasets to a manageable size for regression analysis. Most of the spectral waveband regions selected have been correlated with

key biophysical/biochemical components that have been proven to contribute to yield in many crops. Regression models built based on two-year means explained a large portion of the variability within soybean yield in both maturity groups. The red-edge area and portions of the NIR were the most important portions of the electromagnetic spectra, and creation of yield prediction models based on these waveband regions can account for a significant amount of variability within soybean seed yield. Variability was high within and between growth stages, water regimes, environments, and years for validation of training models. Environmental factors such as weather, time of day, and other factors appear to affect spectral reflectance tremendously, leading the training models performing better or worse on different validation datasets. The performance was not as consistent as hoped, but still accounted for a significant portion of variability in seed yield in most environments. The training models also had the ability to distinguish the top three highest yielding genotypes within the top 50% of the genotypes in most validation datasets.

This experiment demonstrated that canopy reflectance can be used to characterize soybean seed yield using a diverse set of genotypes. These genotypes allowed for significant variation in model training datasets; however, experiments need to be conducted with genotypes that have less diversity to validate the models. Integrating spectral reflectance measurements into a high-throughput platform also is necessary before this technology can be adopted in breeding programs.

References

Adams, M.L., W.D. Philpot, and W.A. Norvell. 1999. Yellowness index: an application of spectral 2nd derivatives to estimate chlorosis of leaves in stressed vegetation. Int. J. of Remote Sensing. 20:3663–3675.

Aparicio, N., D. Villegas, J. Araus, J. Casadesus, and C. Royo. 2002. Relationship between growth traits and spectral vegetation indices in durum wheat. Crop Sci. 42(5):1547–1555.

Babar, M., M. Reynolds, M. van Ginkel, A. Klatt, W. Raun, and M. Stone. 2006a. Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. Crop Sci. 46(2):578–588.

Babar, M., M. Reynolds, M. Van Ginkel, A. Klatt, W. Raun, and M. Stone. 2006b. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. Crop Sci. 46(3):1046–1057.

Ball, S., and C. Kozak. 1993. Relationship between grain-yield and remotely-sensed data in wheat breeding experiments. Plant Breeding. 110(4):277–282.

Baret, F., and G. Guyot. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment, Remote Sens. Environ. 35:161–174.

Blackmer, T.M., J.S. Schepers, and G.E. Varvel. 1994. Light reflectance compared with other nitrogen stress measurements in corn leaves. Agron. J. 86:934–938.

Chang, J., D.E. Clay, K. Dalsted, S. Clay, M. O'Neill. 2003. Corn (*Zea Mays* L.) yield prediction using multispectral and multidate reflectance. Agron. J. 95:1447–1453.

Chappelle, E.W., M.S. Kim, and J.E. McMurtrey, III. 1992. Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. Remote Sens. Environ. 39:239–247.

Clevers, J.G.P.W. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. Remote Sens. Environ. 61:221–228.

Datt, B. 1999. Remote sensing of water content in Eucalyptus leaves. Australian Journal of Botany. 47:909–923.

Datt, B., T.R. McVicar, T.G. Van Niel, D.L.B. Jupp, and J.S. Pearlman. 2003. Pre-processing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. IEEE Transactions on Geoscience and Remote Sensing. 41:1246–1259.

Daughtry, C.S.T., C.L. Walthall, M.S. Kim, E. Brown de Colstoun, and J.E. McMurtrey, III. 2000. Estimating corn leaf chlorophyll content from leaf and canopy reflectance. Remote Sens. Environ. 74:229–239.

De Jong, S. 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems. 18(3):251–263.

Deering, D.W. 1978. Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. Ph.D. diss. Texas A&M Univ., College Station.

Dusek, D.A., R.D. Jackson, and J.t. Musick. 1985. Winter wheat vegetation indices calculated from combinations of seven spectral bands. Remote Sensing of Environment, 18:255–267.

Fehr, W.R., and C.E. Caviness. 1977. Stages of Soybean Development, Cooperative Extension Service, Agriculture and Home Economics Experiment Station Iowa State University, Ames, Iowa

Gamon, J.A., J. Peñuelas, and C.B. Field. 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41:35–44.

Gao, B.C. 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. Sens. Environ. 58:257–266.

Garbulsky, M. 2011. The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies. Remote Sensing of Environment. 115:281–297.

Gitelson, A., Y. Gritz, and M. Merzlyak. 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology. 160:271–282.

Gitelson, A.A., A. Vi.a, D.C. Rundquist, V. Ciganda, and T.J. Arkebauer. 2005. Remote estimation of canopy chlorophyll content in crops. Geophys. Res. Lett. 32:108403 doi:10.1029/2005Gl022688.

Gitelson, A.A., P.S. Thenkabail, J.G. Lyon, A. Huete. 2011. Remote Sensing estimation of crop biophysical characteristics at various scales. Chapter 15 in Hyperspectral Remote Sensing of Vegetation. 329–358, Taylor and Francis.

Gutierrez, M., M. Reynolds, W. Raun, M. Stone, and A. Klatt. 2010. Spectral water indices for assessing yield in elite bread wheat genotypes under well-irrigated, water-stressed, and high-temperature conditions. Crop Sci. 50:197–214.

Gutierrez-Rodriguez, M., M.P. Reynolds, J.A. Escalante-Estrada, M.T. Rodriguez-Gonzalez. 2004. Association between canopy reflectance indices and yield and physiological traits in bread wheat under drought and well-irrigated conditions. Australian J. of Ag. Research. 55: 1139–1147.

Hansen, P.M., J.R. Jorgensen, A. Thomsen. 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J. of Ag. Sci. 139:307–318.

Harshman, R., and M. Lundy. 1994. PARAFAC: Parallel factor analysis. Computational Statistics and Data Analysis, 18:39–72.

Hatfield, J., and J. Prueger. 2010. Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management

practices. Remote Sensing. 2:562–578.

Hatfield, J.L. 1983. Remote sensing estimators of potential and actual crop yield. Remote Sens. Environ. 13:301–311.

Hong, S.Y., K.A. Sudduth, N.R. Kitchen, H.L. Palm, and W.J. Weibold. 2001. Using hyperspectral remote sensing data to quantify within-field spatial variability [CD ROM]. Proc. Int. Conf. on Geospatial Inf. in Agric. and Forestry, 3rd, Denver, CO. 5–7 Nov. 2001. Altarum, Ann Arbor, MI.

Jordan, C.F. 1969. Derivation of leaf area index from quality of light on the forest floor. Ecology 50:663–666.

Kaul, M., R.L. Hill, and C. Walthall. 2005. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems 85:1–18.

Lin, W.S., C.M. Yang, and B.J. Kuo. 2012. Classifying cultivars of rice (*Oryza sativa* L.) based on corrected canopy reflectance spectra data using the orthogonal projections of latent structures (O-PLS) method. Chemometrics and Intelligent Laboratory Systems. 115:25–36.

Loss, S.P., and K.H.M. Siddique. 1994. Morphological and physiological traits associated with wheat yield increases in Mediterranean environments. Advances in Agronomy. 52: 229–276.

Ma, B.L., L.M. Dwyer,C. Costa,E.R. Cober,M.J. Morrison. 2001. Early Prediction of Soybean yield from canopy reflectance measurements. Agron. J. 93:1227–1234.

Marti, J.,J. Bort, G. Slafer, and J. Araus. 2007. Can wheat yield be assessed by early measurements of normalized difference vegetation index? Annals of Applied Biology,

150:253–257.

Morrison, M.J., H.D. Voldeng, and E.R. Cober. 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. Agron. J. 91:685–689.

Naes, T., T. Isaksson, T. Faern, T. Davis. 2004. A User-friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, West Sussex, p. 344.

Peñuelas, J., F. Baret, and I. Filella. 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31:221–230.

Peñuelas, J.,I. Filella, C. Biel,L. Serrano, and R. Save. 1993. The reflectance at the 950–970 mm region as an indicator of plant water status. International Journal of Remote Sensing. 14:1887–1905.

Pimstein, A., A. Karnieli, and D. Bonfil. 2007. Wheat and maize monitoring based on ground spectral measurements and multivariate data analysis. Journal of Applied Remote Sensing. 1:013530.

Pimstein, A., A. Karnieli, S. Bansal, and D. Bonfil. 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. Field Crops Research. 121:125–135.

Prasad, B., B. Carver, M. Stone, M. Babar, W. Raun, and A. Klatt. 2007a. Genetic analysis of indirect selection for winter wheat grain yield using spectral reflectance indices. Crop Sci. 47:1416–1425.

Prasad, B., B. Carver, M. Stone, M. Babar, W. Raun, and A. Klatt,. 2007b. Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under Great Plains conditions. Crop Sci. 47:1426–1440.

Price, J.C., and W.C. Bausch. 1995. Leaf area index estimation from visible and near-infrared reflectance data. Remote Sens of Environ. 52:55–65.

Reynolds, M., Y. Manes.A. Izanloo,P. Langridge. 2009. Phenotyping approaches for physiological breeding and gene discovery in wheat. Annals of Appl. Bio. 155(3):309–320.

Reynolds, M.P., S. Rajaram, and K.D. Sayre. 1999. Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci. 39:1611–1621.

Rouse, J.W., Jr., R.H. Haas, J.A. Schell, and D.W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. p. 309–317. In Proc. Earth Res. Tech. Satellite-1 Symp., Goddard Space Flight Cent., Washington, DC. 10–14 Dec. 1973.

Royo, C., N. Aparicio, D. Villegas, J. Casadesus,P. Monneveux,and J. Araus. 2003. Usefulness of spectral reflectance indices as durum wheat yield predictors under contrasting Mediterranean conditions. International Journal of Remote Sensing. 24:4403–4419.

SAS Institute Inc., SAS 9.2. Cary, NC: SAS Institute Inc., 2008.

Shanahan, J.F., J.S. Schepers, D.D. Francis, G.E. Varvel, W.W. Wilhelm, J.S. Tringe, M.R. Schlemmer, and D.J. Major. 2001. Use of remote sensing imagery to estimate corn grain yield. Agron. J. 93:583–589.

Sims, D.A., AND J.A. Gamon.2003. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. Remote Sens. Environ. 84:526–537.

Stark, P.J., D. Zhao, W.A. Phillips, S.W. Coleman. 2006. Development of canopy reflectance algorithms for real-time prediction of Bermudagrass pasture biomass and nutritive values. Crop Sci. 46: 927–934.

Trotter, G. M., D. Whitehead, and E. J. Pinkney. 2002. The photochemical reflectance index as a measure of photosynthetic light use efficiency for plants with varying foliar nitrogen contents. International J. of Remote Sens. 23:1207−1212.

Tucker, C. J. and P.J. Sellers. 1986. Satellite remote sensing of primary production. International Journal of Remote Sensing. 7:395–1416.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8:127–150.

Voldeng, H.D., E.R. Cober, D.J. Hume, C. Gillard, and M.J. Morrison.1997. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 37:428–431.

Vollmann, J., H. Walter, T. Sato, and P. Schweiger. 2011. Digital image analysis and chlorophyll metering for phenotyping the effects of nodulation in soybean. Computers and Electronics in Agriculture. 75:190–195.

Walter, A., B. Studer, R. Kolliker. 2012. Advanced phenotyping offers opportunities for improved breeding of forage and turf species. Annals of Botany. 1–9.

Weber, V.S., J.L. Araus, J.E. Cairns. C. Sanchez, A.E. Melchinger, E. Orsini. 2012. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. Field Crops Researcch. 128:82–90.

Wenjiang, H., W. Jihua, W. Zhijie, Z. Jiang, L. Liangyun, and W. Jindi. 2004. Inversion of foliar biochemical parameters at various physiological stages and grain quality indicators of winter wheat with canopy reflectance. Int. J. Remote Sensing. 25 (12):2409–2419.

White, J.W., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, M.M. Conley, K.A. Feldmann, A.N. French, J.T. Heun, D.J. Hunsake, M.A. Jenks, B.A. Kimball, R.L. Roth, R.J. Strand, K.R. Thorp, G.W. Wall, and G. Wang. 2012. Field-Based phenomics for plant genetics research. Field Crops Research, 133:101–112.

Wiegand, C., A. Richardson, D Escobar, and A Gerbermann.. 1991. Vegetation indexes in crop assessment. Remote Sensing of Environment. 35:105–119.

Wold, H. 1966. Estimation of principal components and related models by iterative least squares. In Krishnaiaah, P.R.. *Multivariate Analysis*. New York: Academic Press. 391–420.

Zhao, D., K.R. Reddy, V.G. Kakani, J.J. Read, and S. Koti. 2006. Canopy reflectance in cotton for growth assessment and lint yield prediction. Europ. J. Agronomy. 26:335–344.

Table 3.1. Mean seed yield by environments and range in yield across genotypes within environments.

Year	Field	Water regime	Mean	LSD	Range
Maturity group III					
			t ha^{-1}		
2011	A	IRR	2.98	0.43	2.52
2011	A	DRY	1.75	0.60	2.01
2012	B	IRR	3.63	0.58	2.67
2012	B	DRY	2.78	0.92	2.31
2012	C	IRR1	3.99	0.60	2.86
2012	C	IRR2	3.94	0.59	3.33
Maturity group IV					
2011	A	IRR	3.20	0.52	3.27
2011	A	DRY	1.76	0.55	1.81
2012	B	IRR	3.59	0.38	3.63
2012	B	DRY	3.23	0.41	2.49
2012	C	IRR1	3.77	0.74	2.14
2012	C	IRR2	3.99	0.51	2.68

Table 3.2. Maturity group III (MGIII) and maturity group IV (MGIV) analysis of variance F-values for seed yield, maturity (Mat), and spectral wavelengths used for yield estimation models.

Source	DF	Yield	Mat	415	550	680	715	915	940	990	1100	1140	1245	1300
							MGIII							
Gen (G)	19	111.6**	75.0**	10.2**	17.3**	12.6**	14.7**	5.1**	4.5**	4.6**	3.5**	3.9**	4.7**	4.7**
Env (E)	5	109.1**	94.3**	18.3**	11.0**	8.7*	8.3*	45.1**	38.9**	40.8**	36.4**	14.9**	28.3**	23.6**
G × E	95	3.1**	4.5**	3.3**	3.5**	3.2**	3.5**	2.1*	2.0*	1.9*	1.7*	1.5*	1.7*	1.7*
							MGIV							
Gen (G)	19	99.9**	40.4**	90.9**	16.3**	8.5**	14.9**	3.0*	2.5*	2.8*	2.9*	2.2*	2.1*	2.1*
Env (E)	5	84.5**	567.4**	112.3**	137.6**	139.7**	139.6**	50.7**	59.3**	55.5**	49.5**	64.9**	61.1**	58.8**
G × E	95	3.9**	2.6**	1.0	1.4*	1.1	1.5*	0.8	0.6	0.6	0.6	0.5	0.5	0.5

* = Pr < 0.05 ** = Pr < 0.01

Table 3.3. Maturity group III (MGIII) and maturity group IV (MGIV) analysis of variance F-values for 2011 and 2012 experiments, and wavelengths used for yield estimation models.

MGIII

				Wavelength (nm)										
Year	Field-Env	Growth Stage	DF	415	550	680	715	915	940	990	1100	1140	1245	1300
2011	A-I	R3-R4	19	2.77**	3.41**	2.20**	2.94**	1.57	1.58	1.43	0.96	1.29	0.74	0.79
2011	A-I	R4	19	6.27**	9.82**	4.36**	9.98**	3.29**	2.86**	2.46**	2.32**	2.08**	2.25**	2.27**
2011	A-I	R5	19	3.77**	8.22**	1.86*	5.55**	3.71**	3.10**	2.41**	1.96*	1.80*	1.84*	1.86*
2011	A-I	R6	19	3.15**	8.31**	8.19**	5.73**	3.86**	3.48**	2.94**	2.24**	2.02**	1.93*	1.87*
2011	A-D	R3-R4	19	4.31**	6.89**	3.35**	6.79**	1.84*	1.75*	1.83*	1.35	1.44	1.52	1.52
2011	A-D	R5	19	7.16**	19.44**	5.92**	17.97**	4.26**	3.75**	3.97**	3.58**	3.13**	3.93**	3.93**
2011	A-D	R6	19	16.12**	17.37**	18.58**	10.90**	8.31**	7.06**	5.61**	4.12**	2.79**	2.12**	1.95*
2012	B-I	R2	19	5.35**	8.81**	3.93**	9.17**	1.31	3.49**	2.47**	3.18**	2.33**	3.33**	3.30**
2012	B-I	R3	19	2.73**	4.32**	2.38**	4.57**	1.44	1.43	1.45	NS	1.42	1.50	1.51
2012	B-I	R3-R4	19	2.03**	3.84**	2.15**	4.12**	1.71*	1.70*	1.78*	1.59	1.69*	1.82*	1.83*
2012	B-I	R5-R6	19	6.16**	8.86**	7.00**	9.53**	1.82*	1.98*	2.24**	2.07**	2.49**	2.77**	2.84**
2012	B-D	R3	19	3.35**	2.95**	1.97*	4.14**	1.67*	1.66*	1.65	1.30	1.40	1.64	1.67*
2012	B-D	R4	19	6.64**	7.85**	5.47**	11.26**	1.91*	1.72*	1.85*	1.64	1.72*	2.22**	2.31**
2012	C-I 1	R2	19	9.38**	18.06**	9.25**	13.45**	1.45	1.60	1.67*	1.39	1.77*	1.92*	2.01**
2012	C-I 1	R3	19	1.71*	3.34**	1.94*	3.01**	0.73	0.71	0.71	NS	0.72	0.74	0.75
2012	C-I 1	R4-R5	19	3.38**	4.50**	2.49**	5.21**	1.61	1.76*	1.77*	1.46	1.80*	1.85*	1.92*
2012	C-I 1	R6	19	5.58**	6.01**	4.44**	7.63**	2.75**	2.65**	3.34**	2.72**	2.80**	3.65**	3.69**
2012	C-I 2	R2	19	1.83*	2.73**	1.61*	3.93**	1.45	1.46	1.54	1.70*	1.38	1.56	1.59
2012	C-I 2	R4-R5	19	4.17**	6.29**	3.69**	9.74**	4.01**	4.29**	4.54**	3.00**	3.74**	4.09**	4.14**

MGIV

				Wavelength (nm)										
Year	Field-Env	Growth Stage	DF	415	550	680	715	915	940	990	1100	1140	1245	1300
2011	A-I	R3	19	2.87**	3.91**	2.35**	4.57**	2.20**	2.09**	1.59	1.57	1.52	1.43	1.45
2011	A-I	R3-R4	19	3.92**	6.46**	4.38**	4.81**	4.93**	4.09**	3.50**	3.10**	2.11**	2.11**	2.00*
2011	A-I	R5	19	4.75**	6.59**	3.87**	4.94**	9.78**	8.28**	6.27**	5.03**	3.70**	2.84**	2.68**
2011	A-I	R5-R6	19	2.81**	3.01**	2.81**	2.65**	5.57**	5.00**	3.51**	2.64**	2.14**	1.69*	1.64
2011	A-D	R2-R3	19	7.05**	6.57**	4.98**	5.11**	2.13**	2.01**	2.00**	1.98*	1.77*	1.83*	1.80*
2011	A-D	R3-R4	19	7.87**	7.36**	4.25**	6.19**	4.05**	4.03**	3.73**	3.50**	3.43**	3.36**	3.34**
2011	A-D	R5	19	6.17**	5.69**	6.41**	6.69**	7.84**	7.22**	5.42**	4.28**	3.23**	2.45**	2.36**
2012	B-I	R1-R2	19	5.01**	8.20**	3.02**	9.89**	4.52**	4.93**	3.94**	3.39**	3.54**	2.95**	2.97**
2012	B-I	R2-R3	19	3.91**	4.77**	2.37**	4.23**	0.76	1.03	1.23	1.43	1.95*	1.75*	1.48
2012	B-I	R4	19	1.82*	2.30**	1.84*	2.32**	1.68	1.78*	1.66	2.00*	1.68*	1.29	1.32
2012	B-I	R5	19	4.58**	6.97**	4.08**	6.77**	3.05**	2.77**	2.77**	2.16**	1.84*	2.02**	1.98*
2012	B-I	R5-R6	19	3.09**	4.74**	2.96**	4.32**	2.81**	2.62**	2.47**	2.45**	2.31**	2.34**	2.34**
2012	B-D	R3	19	4.63**	3.43**	2.33**	3.49**	1.72*	1.88*	1.76*	1.36	1.48	1.51	1.75*
2012	B-D	R4	19	5.58**	4.97**	3.34**	4.57**	3.96**	3.45**	3.22**	3.26**	2.58**	2.61**	2.54**
2012	B-D	R6	19	3.81**	6.89**	3.47**	6.85**	6.35**	5.74**	4.89**	4.69**	3.90**	3.48**	3.37**

2012	C-I	R1-R2	19	7.04**	11.90**	6.55**	8.27**	1.10	1.05	1.33	1.24	1.28	1.58	1.58
2012	C-I 1	R2-R3	19	1.67*	2.17**	1.05	1.69*	0.77	0.71	0.72	0.65	0.78	0.60	0.50
2012	C-I 1	R4	19	4.99**	6.19**	3.57**	5.98**	2.57**	2.41**	2.35**	1.91*	1.77*	1.82	1.78*
2012	C-I 1	R5-R6	19	8.10*	7.30**	6.39**	7.71**	3.27**	3.40**	3.26**	2.82**	2.91**	2.78**	2.77**
2012	C-I 2	R3	19	1.53	1.33	1.44	1.67*	1.21	1.29	2.07**	1.08	1.78*	2.02**	2.05**
2012	C-I 2	R4	19	1.37	1.31	1.26	1.29	1.89*	1.83*	1.82*	1.58	1.49	1.51	1.47

* = Pr>0.05, ** = Pr>0.01

Table 3.4. Maturity group III (MGIII) Pearson's correlation coefficients and p-values for two-year averages of seed yield, maturity, and wavebands used for yield estimation models. MGIII on upper right; maturity group IV (MGIV) on lower left.

	Yield	mat	415	550	680	715	915	940	990	1100	1140	1245	1300
yield	-----	0.35	-0.58**	-0.80**	-0.64**	-0.83**	0.39	0.30	0.18	0.27	0.06	-0.12	-0.17
mat	0.68**	-----	0.19	-0.12	0.01	-0.18	0.72**	0.73**	0.70**	0.69**	0.66**	0.53*	0.51*
415	-0.60**	-0.48*	-----	0.84**	0.93**	0.77**	0.04	0.16	0.26	0.20	0.37	0.47*	0.50*
550	-0.75**	-0.52*	0.85**	-----	0.93**	0.98**	-0.08	0.02	0.14	0.07	0.27	0.43	0.48*
680	-0.78**	-0.59**	0.91**	0.94**	-----	0.87**	-0.01	0.10	0.20	0.15	0.31	0.44*	0.48*
715	-0.75**	-0.53*	0.78**	0.99**	0.91**	-----	-0.07	0.03	0.16	0.07	0.28	0.46*	0.51*
915	0.46*	0.57**	-0.17	0.03	-0.17	0.08	-----	0.99**	0.96**	0.98**	0.92**	0.82**	0.80**
940	0.39	0.53*	-0.12	0.10	-0.09	0.16	0.99**	-----	0.99**	0.99**	0.96**	0.88**	0.86**
990	0.19	0.36	0.07	0.31	0.13	0.38	0.94**	0.97**	-----	0.99**	0.99**	0.94**	0.92**
1100	0.32	0.46*	0.01	0.21	0.05	0.26	0.96**	0.97**	0.97**	-----	0.96**	0.90**	0.88**
1140	0.03	0.21	0.19	0.46*	0.30	0.53*	0.85**	0.90**	0.97**	0.93**	-----	0.97**	0.96**
1245	-0.12	0.05	0.34	0.60**	0.47*	0.66**	0.73**	0.79**	0.91**	0.85**	0.98**	-----	1.00**
1300	-0.18	0.01	0.37	0.64**	0.51*	0.70**	0.69**	0.76**	0.89**	0.82**	0.96**	1.00**	-----

* Pr > 0.05 ** Pr > 0.01

Table 3.5. Results of the maturity group III (MGIII) stepwise regression models by growth stage within environments, coefficient of determination (R2), root means square error (rMSE), percentage rMSE of dependent means (% rMSE of mean), and waveband(s) in final model.

Environment	Growth stage	R^2	rMSE t ha^{-1}	% rMSE of mean	Waveband(s)
A-I	**R3-R4**	**0.73**	**0.47**	**15.66**	**415 680**
A-I	**R4**	**0.66**	**0.53**	**17.64**	**715 940**
A-I	**R5**	**0.62**	**0.56**	**18.80**	**550 915**
A-I	**R6**	**0.69**	**0.51**	**17.02**	**415 715**
A-D	R3-R4	0.51	0.49	26.99	680 940
A-D	R5	0.57	0.46	25.35	715 940
A-D	R6	0.55	0.45	25.08	680
B-D	**R3**	**0.66**	**0.42**	**14.49**	**715 915**
B-D	**R4**	**0.70**	**0.38**	**13.32**	**715**
B-I	R2	0.47	0.66	17.18	550
B-I	R3	NS	NS	NS	
B-I	R3-R4	NS	NS	NS	
B-I	**R5-R6**	**0.80**	**0.44**	**11.86**	**550 715 915**
C-I 2	**R2**	**0.66**	**0.51**	**12.47**	**550 715 915**
C-I 2	**R4-5**	**0.79**	**0.40**	**10.02**	**415 550 915 940**
C-I 1	**R2**	**0.78**	**0.37**	**9.10**	**680 715**
C-I 1	R3	NS	NS	NS	
C-I 1	**R4-5**	**0.78**	**0.38**	**9.30**	**550 715 990**
C-I 1	**R6**	**0.85**	**0.32**	**7.79**	**550 715 990**

BOLD = Selected observations used for training model creation.

Table 3.6. Results of the maturity group IV (MGIV) stepwise regression models by growth stage with experiment, coefficient of determination (R2), root means square error (rMSE), percentage rMSE of dependent mean (% rMSE of mean), and wavebands in final model (wavebands).

Experiment	Growth stage	R^2	rMSE t ha^{-1}	% rMSE of mean	Wavebands
A-I	**R3**	**0.73**	**0.56**	**17.10**	**715 1100**
A-I	**R3-R4**	**0.89**	**0.39**	**11.86**	**415 680 715 915 1300**
A-I	**R5**	**0.75**	**0.53**	**16.45**	**715 915**
A-I	**R5-R6**	**0.73**	**0.56**	**17.27**	**715 915**
A-D	**R2-R3**	**0.69**	**0.31**	**17.48**	**940 1100 1245**
A-D	R3-R4	0.54	**0.38**	21.25	680 715 940
A-D	**R5**	**0.66**	**0.31**	**17.14**	**550**
B-D	**R3**	**0.79**	**0.37**	**11.32**	**715 940**
B-D	**R4**	**0.77**	**0.39**	**11.78**	**715 1100**
B-D	**R6**	**0.67**	**0.45**	**13.69**	**550**
B-I	**R1-R2**	**0.79**	**0.47**	**12.94**	**715 940**
B-I	R2-R3	NS	NS	NS	
B-I	R4	0.57	0.66	18.38	715 915
B-I	**R5**	**0.86**	**0.40**	**11.07**	**550 715 915 990**
B-I	**R5-R6**	**0.71**	**0.53**	**14.62**	**550**
C-I 2	R3	NS	NS	NS	
C-I 2	R4	NS	NS	NS	
C-I 1	**R1-R2**	**0.52**	**0.43**	**11.15**	**715**
C-I 1	R2-R3	0.47	0.45	11.74	715
C-I 1	R4	0.30	0.52	13.49	415
C-I 1	R5-R6	0.37	0.49	12.80	415

BOLD = Selected observations used for training model creation.

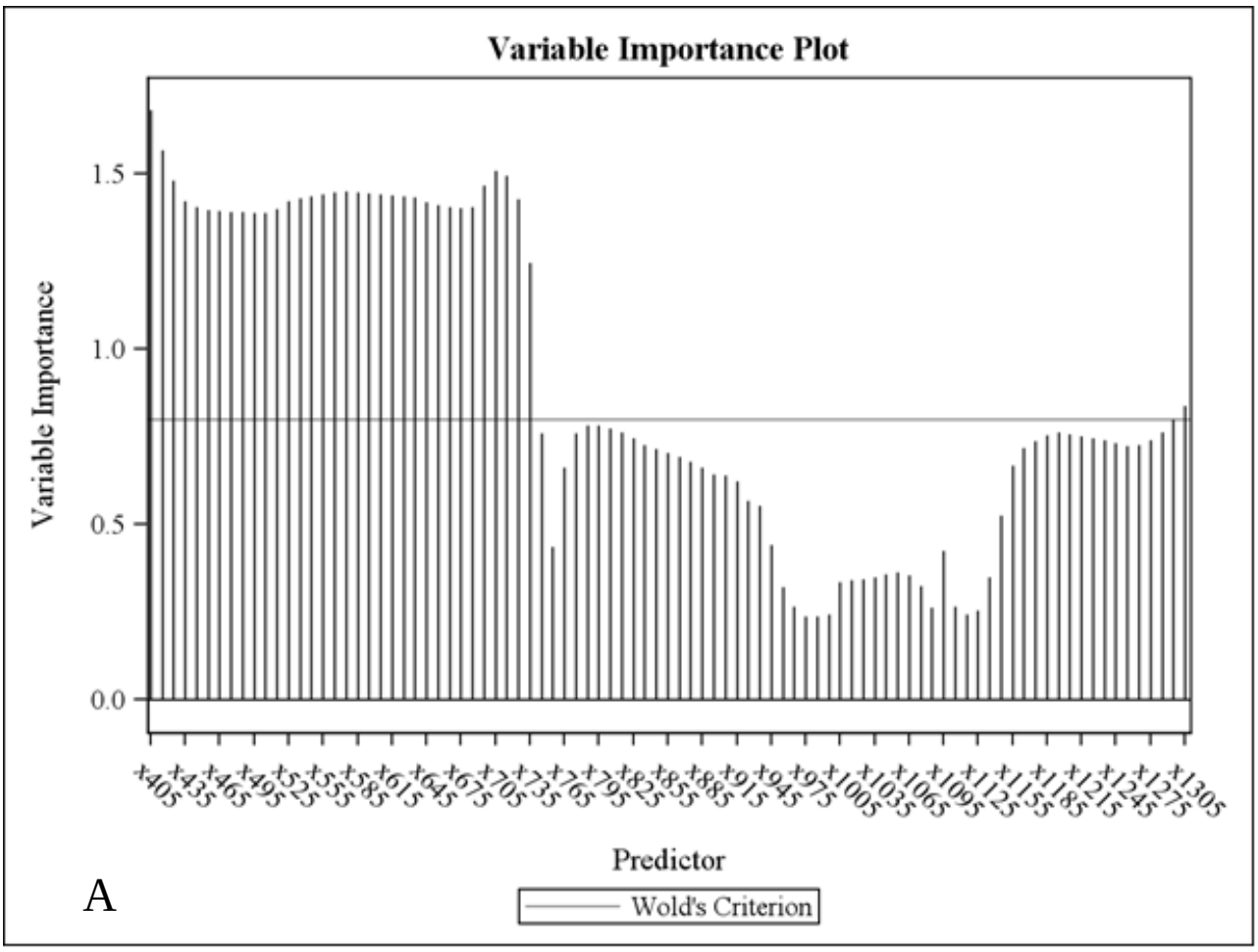
Table 3.7. Results of stepwise regression yield estimation models for maturity group III (MGIII) and maturity group IV (MGIV) selected datasets.

Model	Variable (nm)	Variable R^2	Model R^2	Equation
MGIII	715	0.73	0.83	Yield = -52.35 - 556.42 (715 nm) + 275.28 (1100 nm)
	1100	0.10		
MGIV	715	0.70	0.81	Yield = -45.13 - 619.09 (715 nm) + 172.24 (915 nm)
	915	0.11		

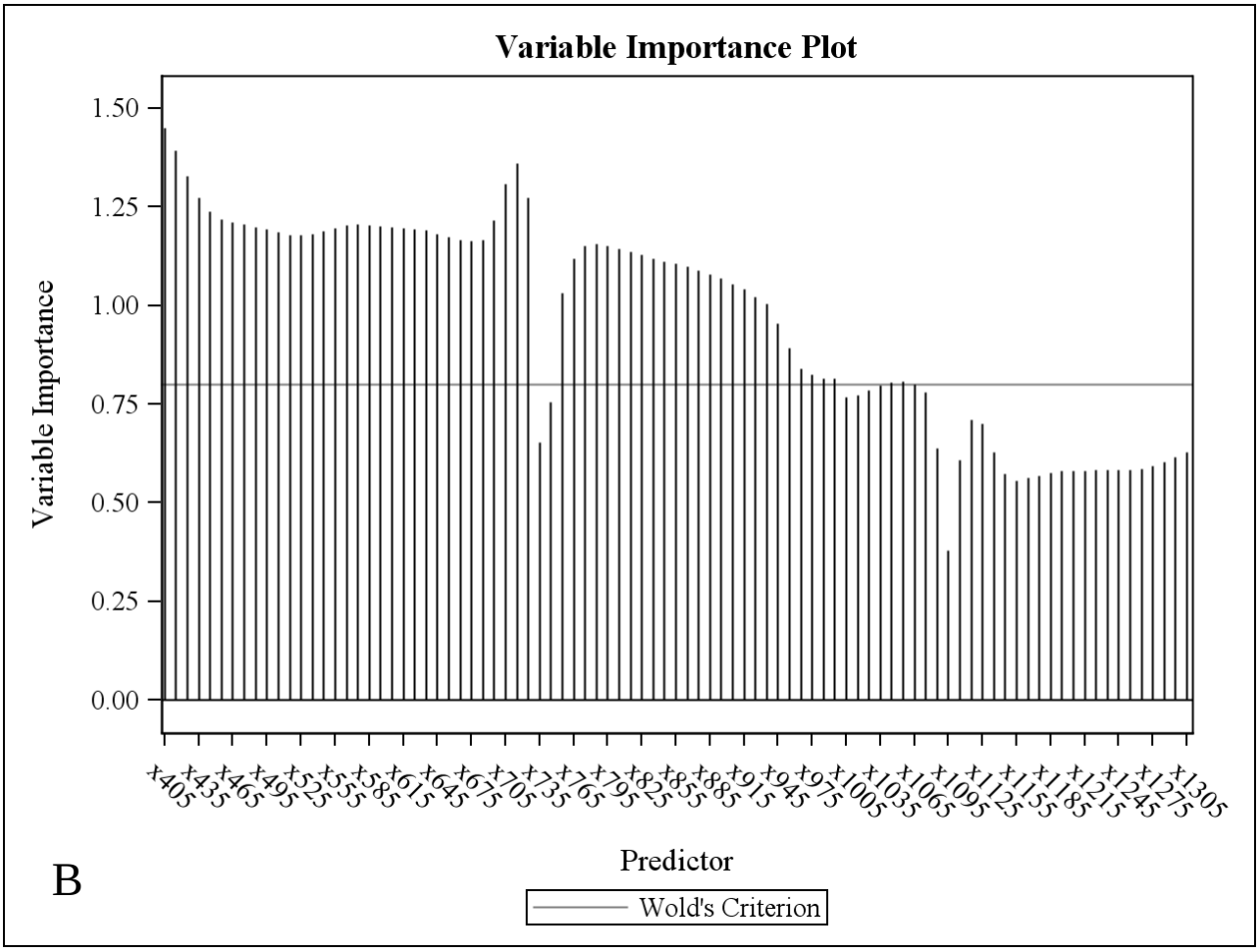
Table 3.8. Yield estimation model validation results for maturity group III (MGIII) and maturity group IV (MGIV) growth stages and season totals (ST) for optimized yield

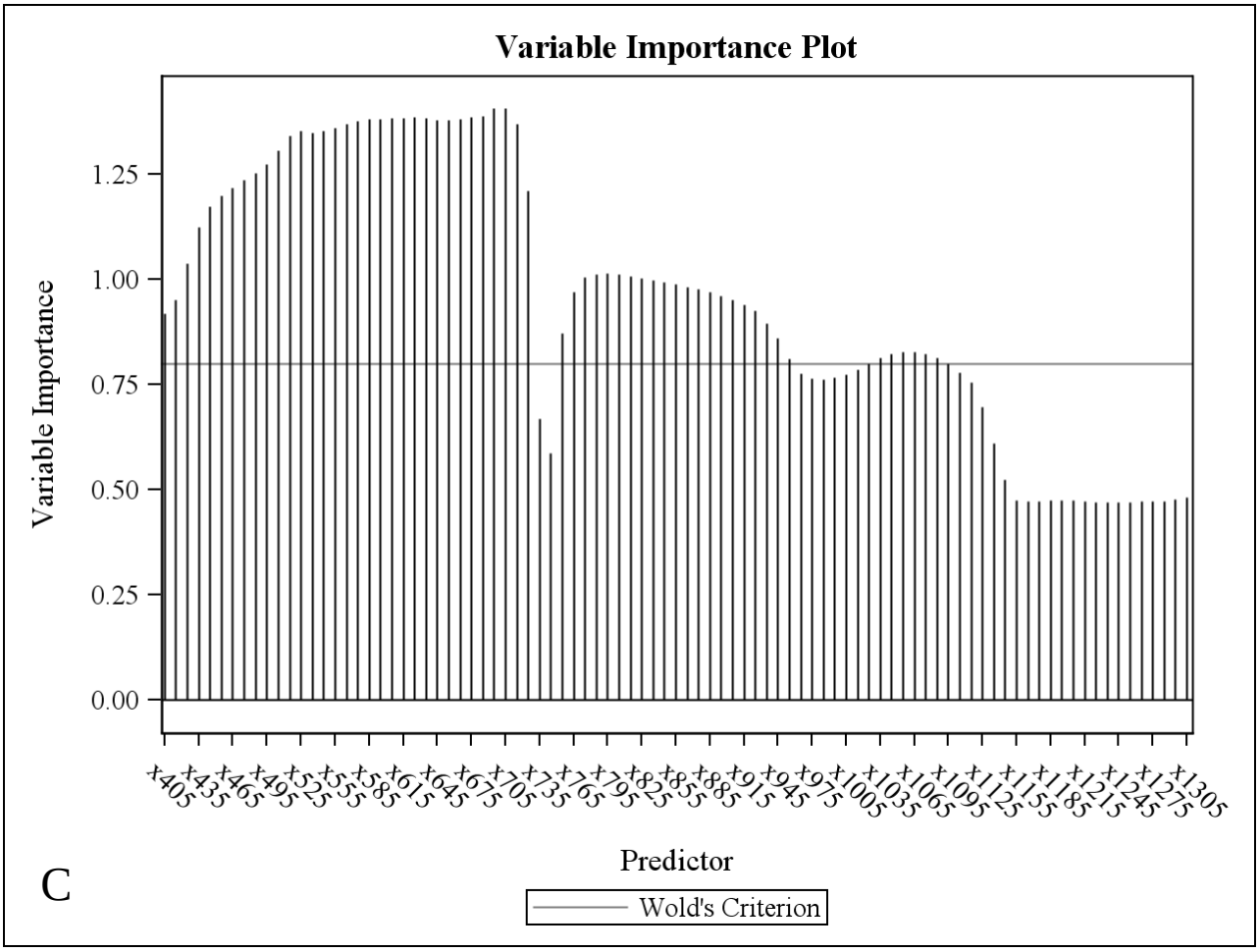
estimation models. I indicates an irrigated environment; D indicates a water-stressed environment.

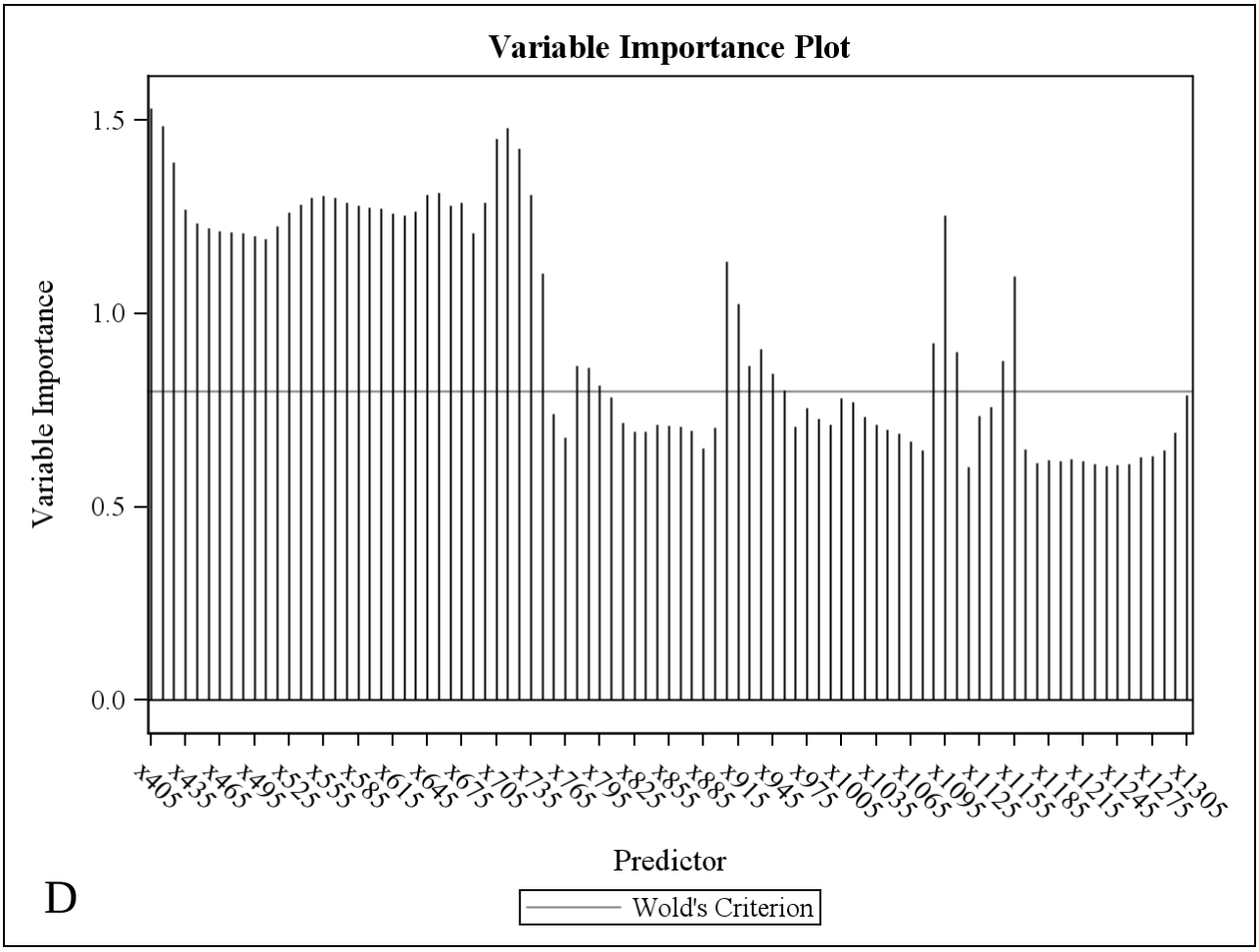
Environment	Growth stage	R^2	Environment	Growth stage	R^2
MGIII			MGIV		
A-I	R3-R4	0.49	A-I	R3	0.72
A-I	R4	0.51	A-I	R3-R4	0.75
A-I	R5	0.66	A-I	R5	0.75
A-I	R6	0.62	A-I	R5-R6	0.72
A-I	ST	0.53	A-I	ST	0.78
A-D	R3-R4	0.62	A-D	R2-R3	0.43
A-D	R5	0.55	A-D	R3-R4	0.44
A-D	R6	0.64	A-D	R5	0.60
A-D	ST	0.67	A-D	ST	0.51
B-D	R3	0.64	B-D	R3	0.77
B-D	R4	0.61	B-D	R4	0.72
B-D	ST	0.66	B-D	R6	0.64
B-I	R2	0.54	B-D	ST	0.83
B-I	R3-R4	0.29	B-I	R1-R2	0.79
B-I	R5-R6	0.64	B-I	R4	0.38
B-I	ST	0.59	B-I	R5	0.56
C-I 2	R2	0.46	B-I	R5-R6	0.61
C-I 2	R4-5	0.42	B-I	ST	0.68
C-I 2	ST	0.45	C-I 2	R3	0.01
C-I 1	R2	0.75	C-I 2	R4	0.01
C-I 1	R4-5	0.53	C-I 2	ST	0.02
C-I 1	R6	0.64	C-I 1	R1-R2	0.54
C-I 1	ST	0.79	C-I 1	R2-R3	0.41
2011 avg.	ST	0.68	C-I 1	R4	0.28
2012 avg.	ST	0.68	C-I 1	R5-R6	0.40
			C-I 1	ST	0.49
			2011 avg.	ST	0.65
			2012 avg.	ST	0.79


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion

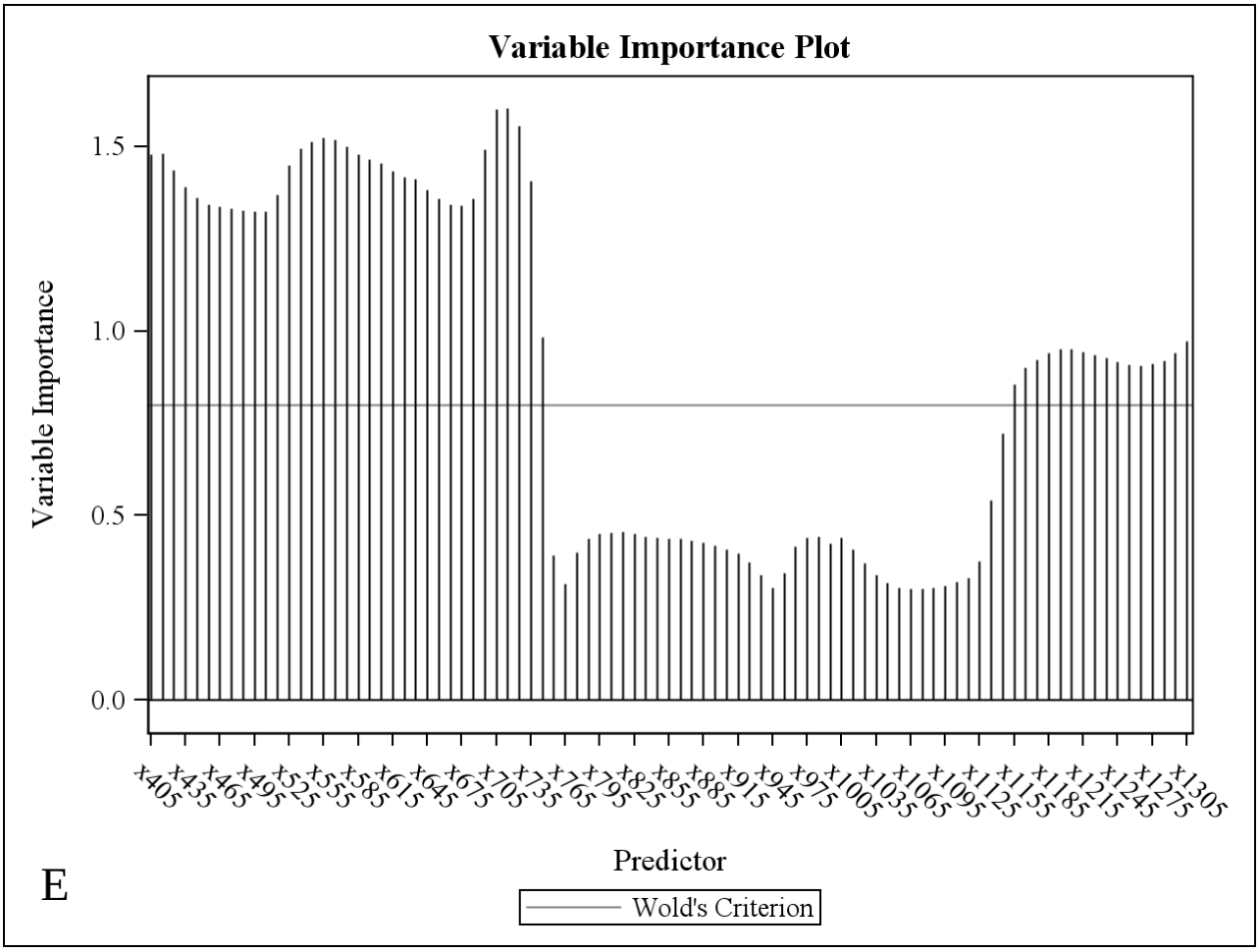
A

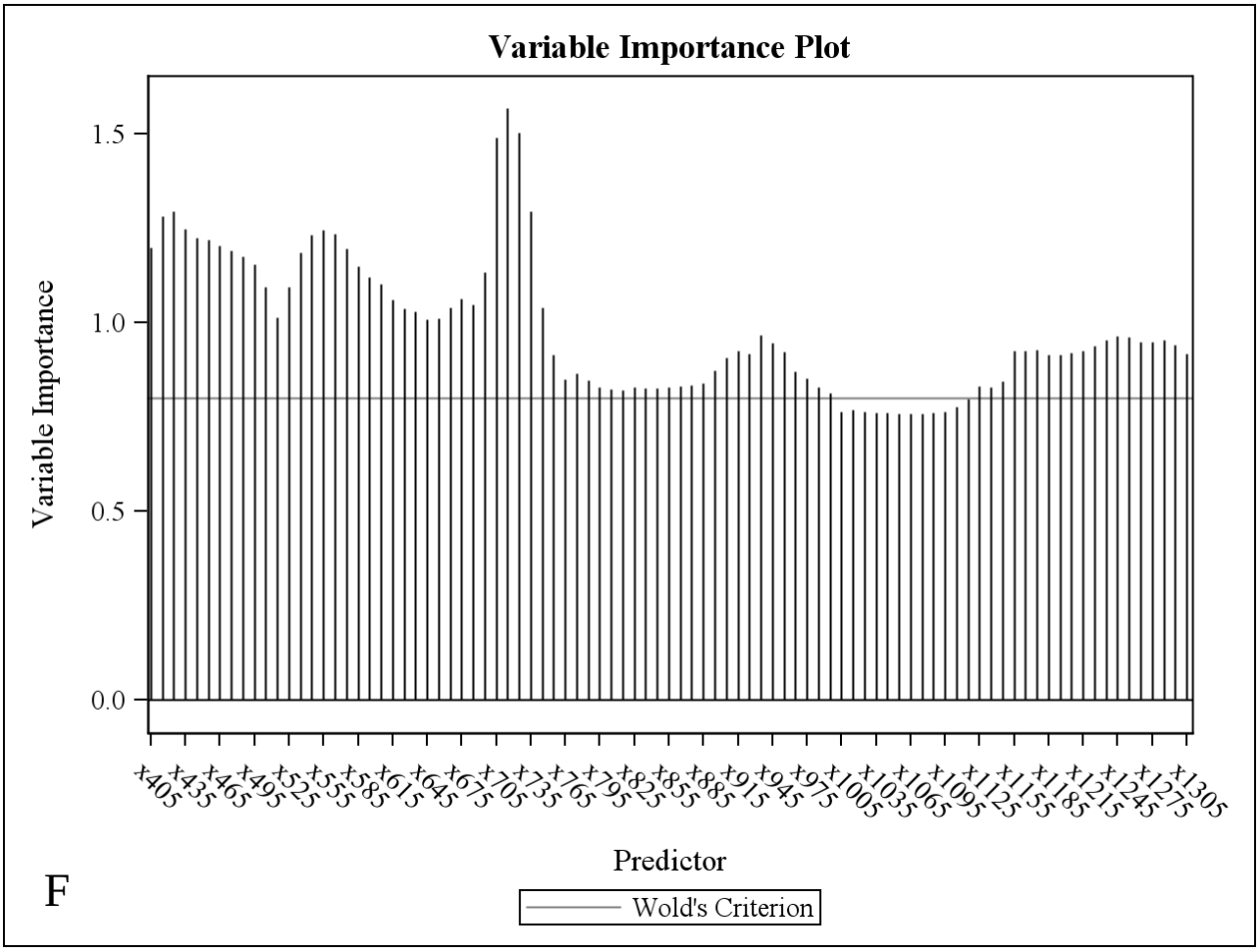

Variable Importance Plot
Variable Importance
1.50
1.25
1.00
0.75
0.50
0.25
0.00
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
B
Wold's Criterion


Variable Importance Plot
Variable Importance
1.25
1.00
0.75
0.50
0.25
0.00
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion
C

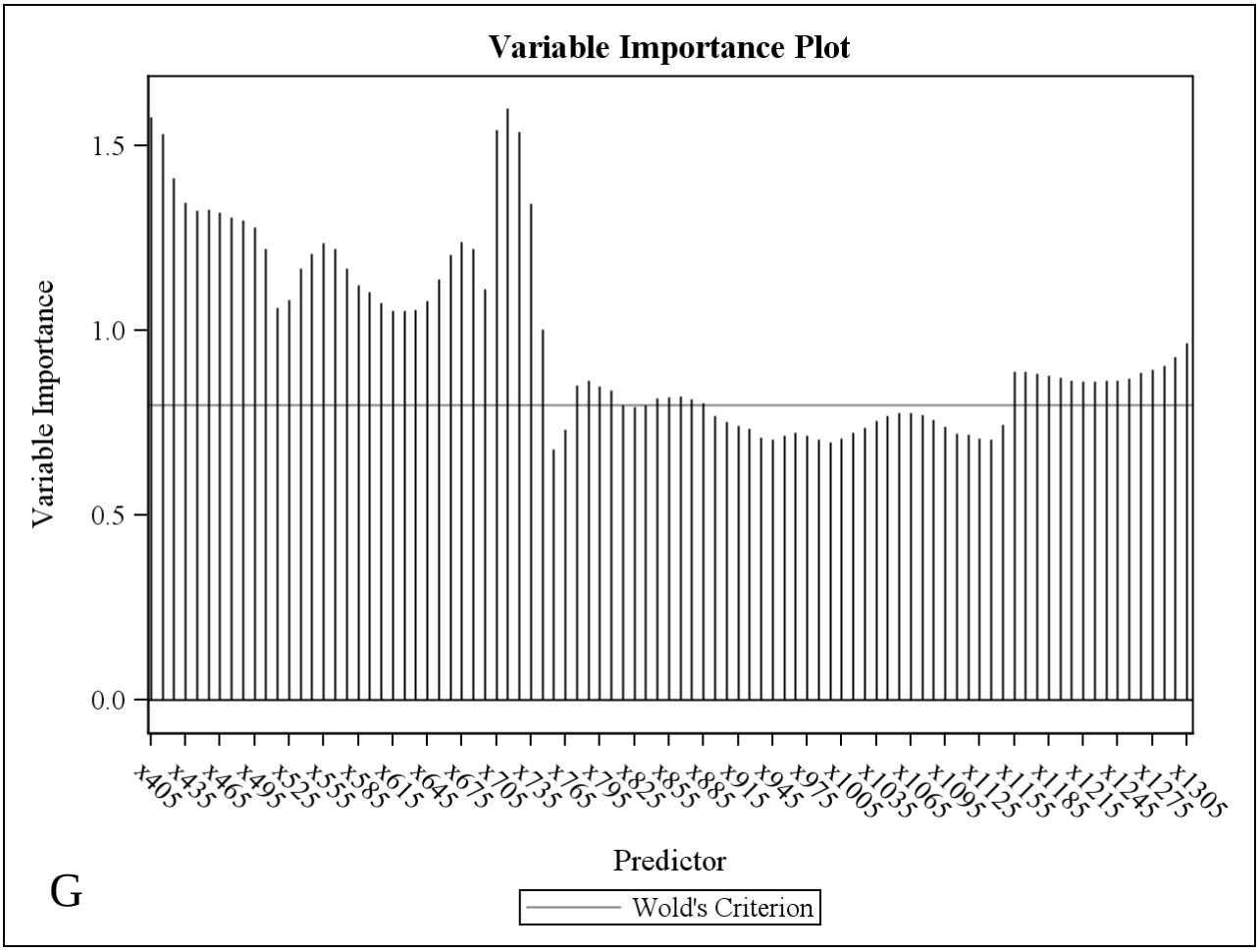

Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion

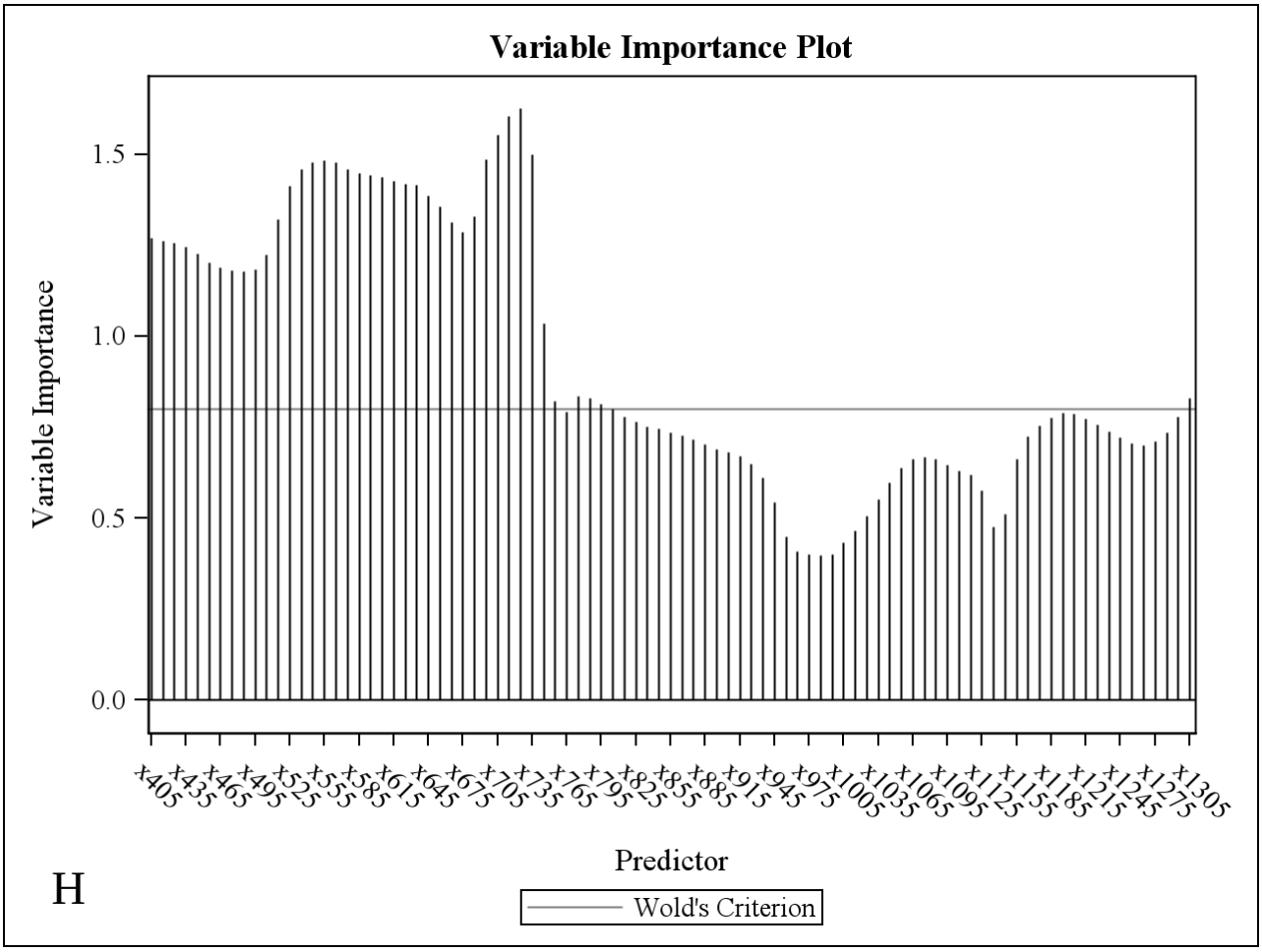
D


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion
E

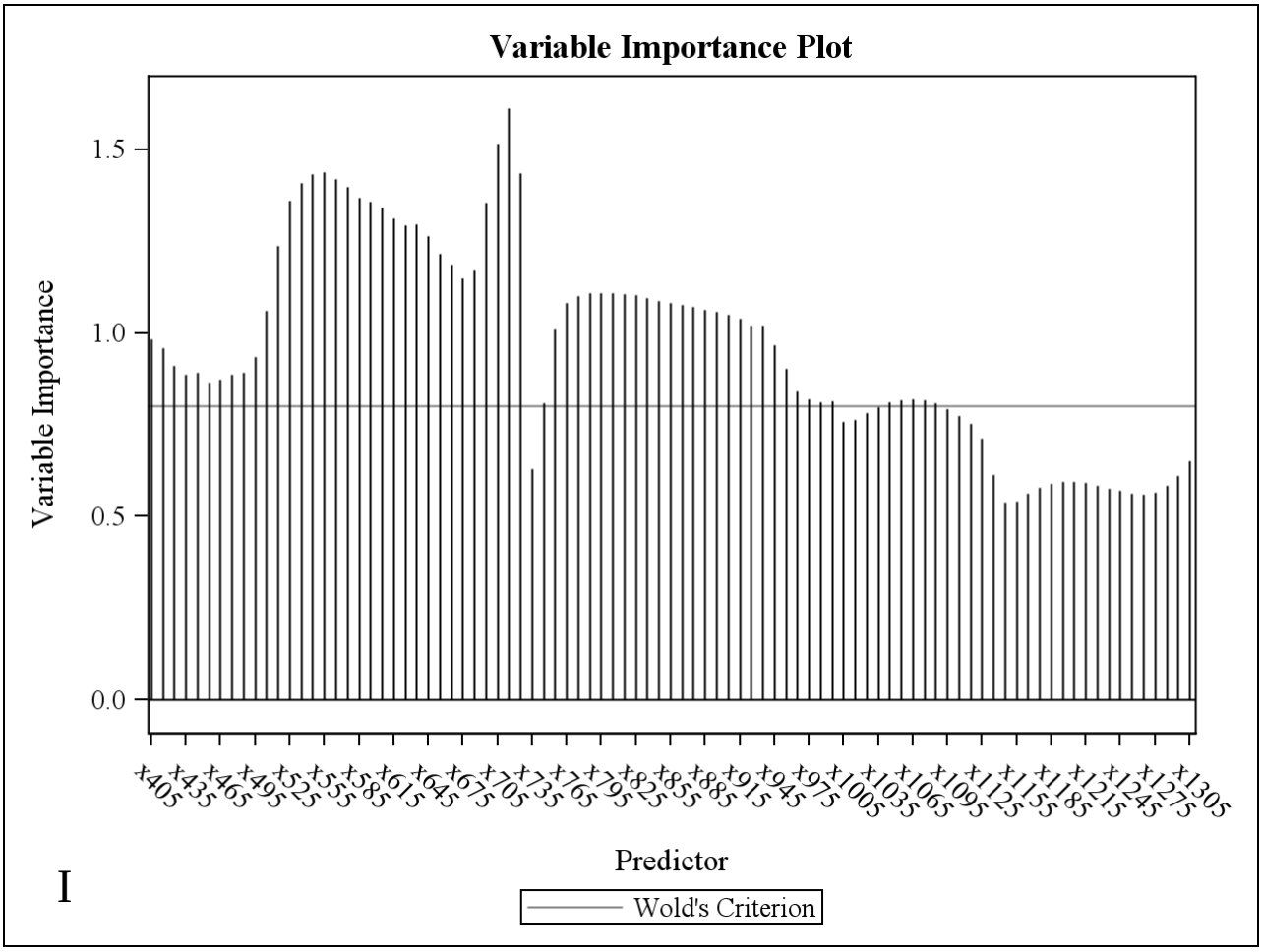

Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion

F

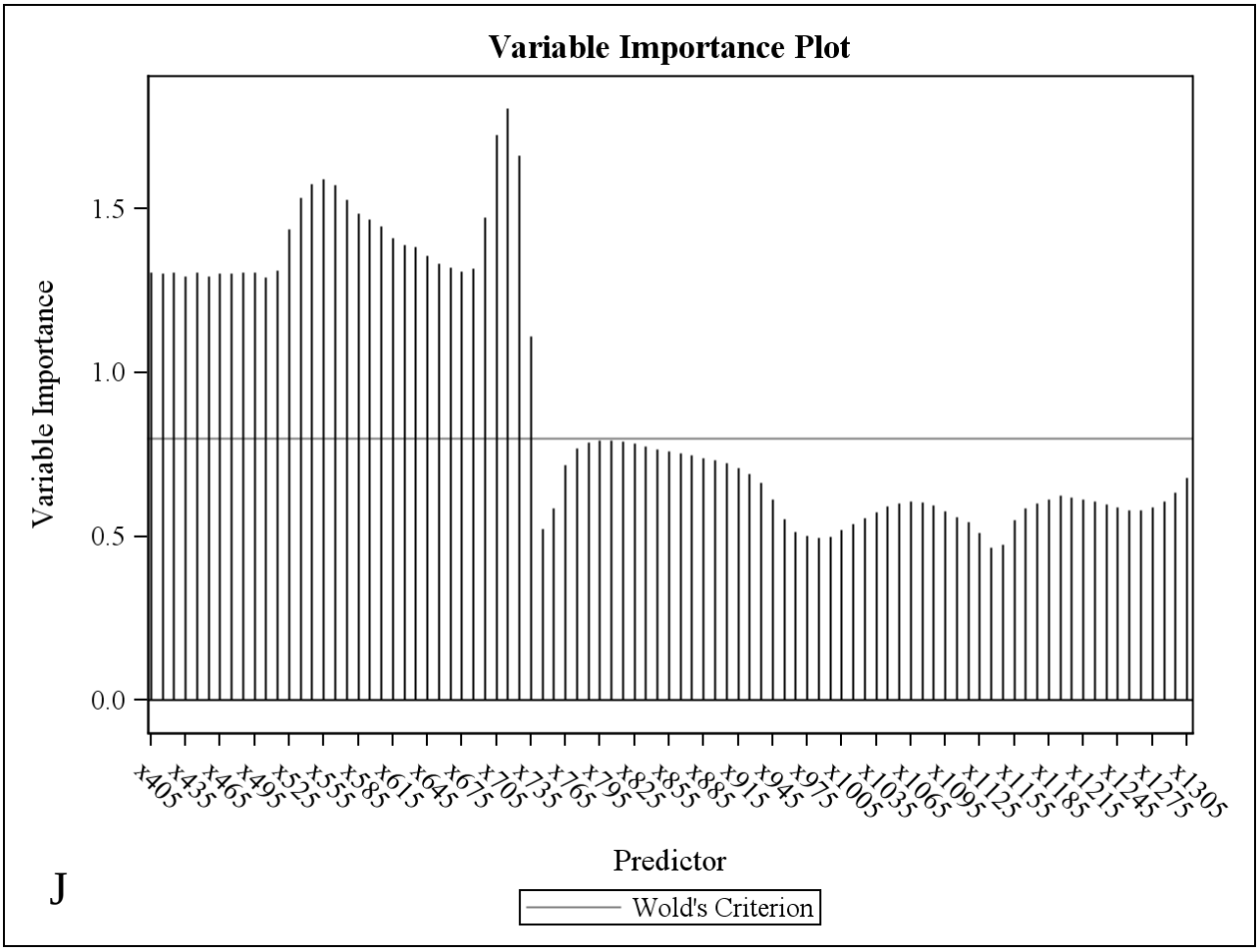

Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
G
Wold's Criterion

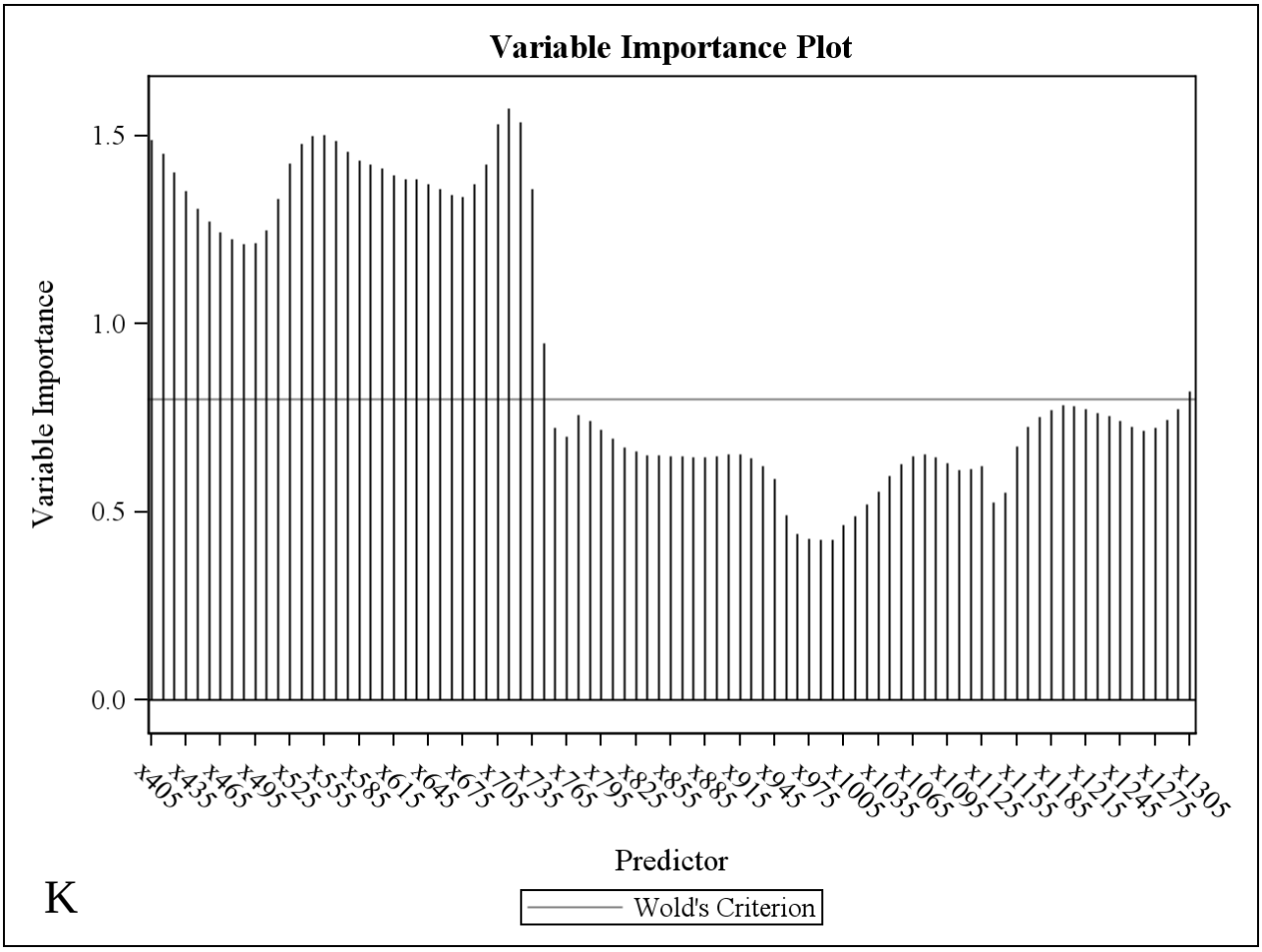

Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion

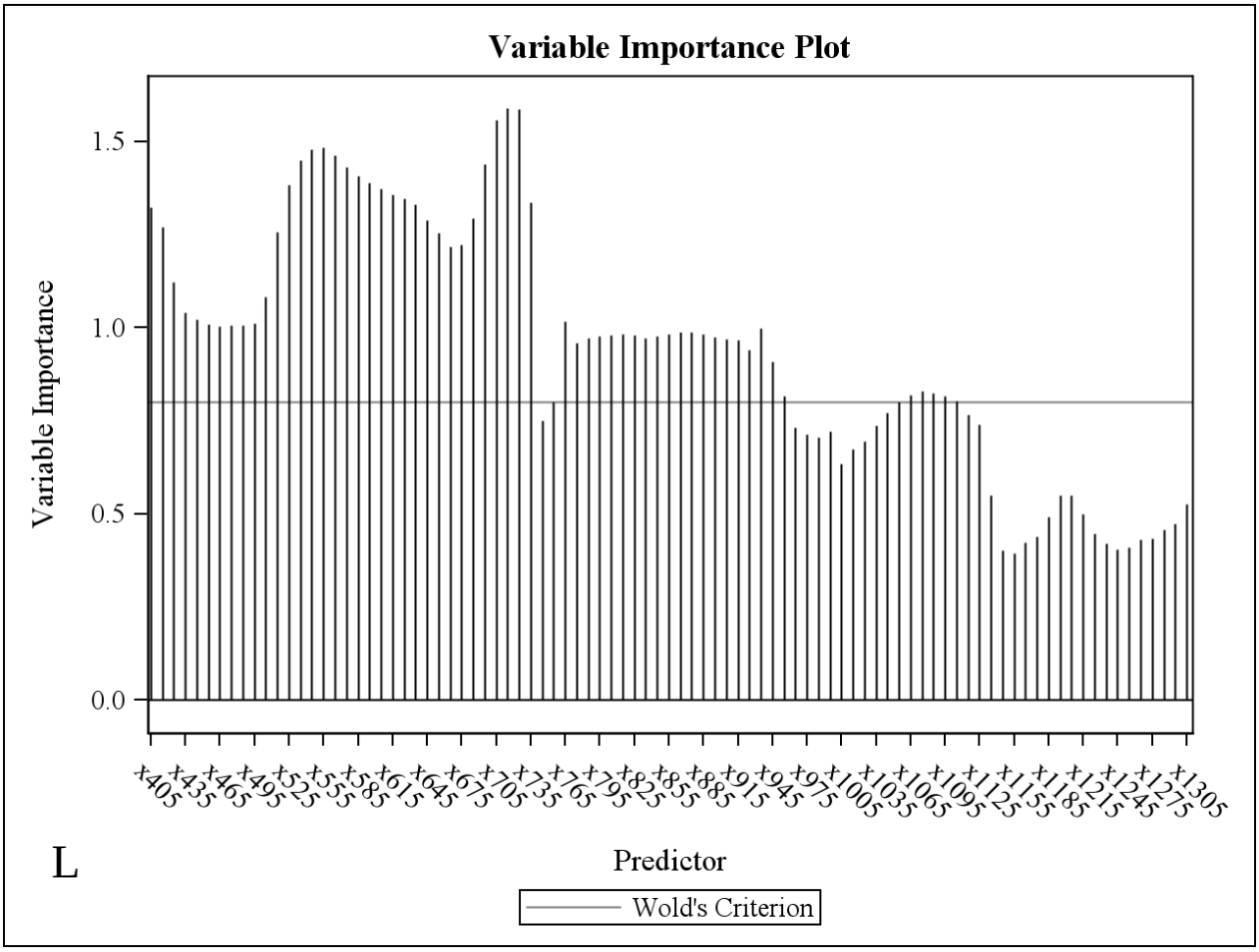
H


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion

I


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
J
Wold's Criterion


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
Predictor
Wold's Criterion
K


Variable Importance Plot
Variable Importance
1.5
1.0
0.5
0.0
x405
x435
x465
x495
x525
x555
x585
x615
x645
x675
x705
x735
x765
x795
x825
x855
x885
x915
x945
x975
x1005
x1035
x1065
x1095
x1125
x1155
x1185
x1215
x1245
x1275
x1305
L
Predictor
Wold's Criterion

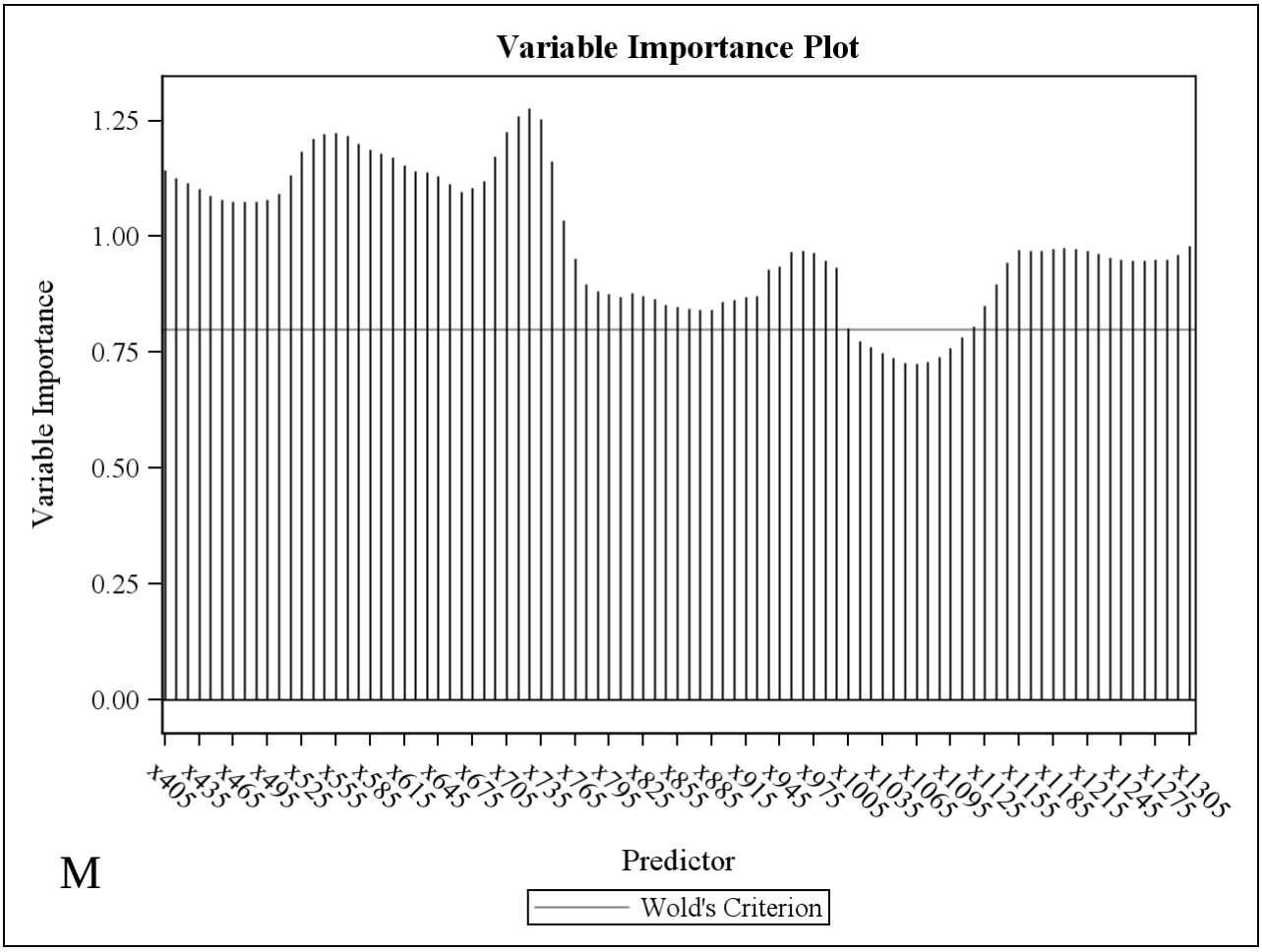


Figure 3.A. Variable importance in projection plots displaying results from partial least squares regression (PLS) analyses. (A) Maturity group III (MGIII) 2011–2012 two-year means, (B) MGIII 2011 A-I means, (C) MGIII 2011 A-D means, (D) MGIII 2012 B-I means, (E) MGIII 2012 B-D means, (F) MGIII 2012 C-I 1 means, (G) MGIII 2012 C-I 2 means, (H) maturity group IV (MGIV) 2011–2012 two-year means, (I) MGIV 2011 A-I means, (J) MGIV 2011 A-D means, (K) MGIV 2012 B-I means, (L) MGIV 2012 B-D means, (M) MGIV 2012 C-I 1 means.

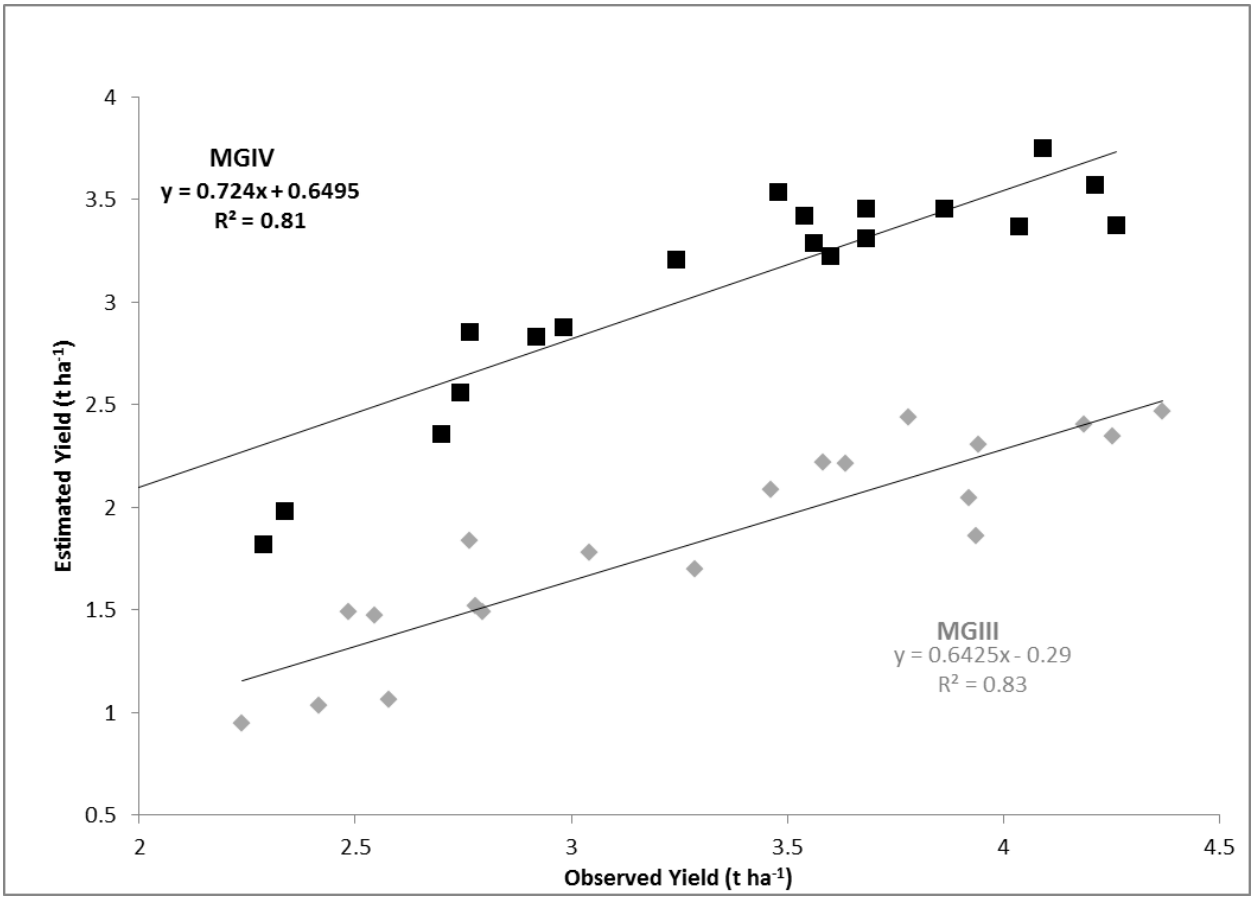


Figure 3.B. Relationships between observed and predicted seed yield for maturity group III (MGIII) and maturity group IV (MGIV) two-year waveband means with selected growth stage